PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7:

C12N 15/54, 9/10, 15/81, 15/82, 1/16, 5/10, A01N 27/067, C12P 7/64

(11) International Publication Number:

WO 00/60095

(43) International Publication Date:

12 October 2000 (12,10,00)

(21) International Application Number:

PCT/EP00/02701

A₂

(22) International Filing Date:

28 March 2000 (28.03.00)

(30) Priority Data:

99106656.4 1 April 1999 (01.04.99) EP 99111321.8 10 June 1999 (10.06.99) EP 60/180,687 7 February 2000 (07.02.00) US

(71) Applicant (for all designated States except US): BASF PLANT SCIENCE GMBH [DE/DE]; D-67056 Ludwigshafen (DE).

(72) Inventors; and

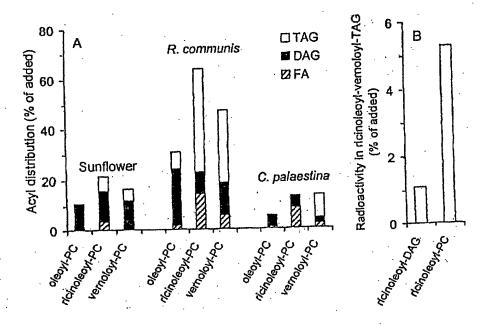
- (75) Inventors/Applicants (for US only): DAHLQVIST, Anders [SE/SE]; Hemmansvägen 2, S-244 66 Furulund (SE). STAHL, Ulf [SE/SE]; Liljegatan 7b, S-753 24 Uppsala (SE). LENMAN, Marit [SE/SE]; Revingegatan 13a, S-223 59 Lund (SE). BANAS, Antoni [SE/PL]; Wiolinowa 14, PL-08110 Siedlce (PL). RONNE, Hans [SE/SE]; Dirigentvägen 169, S-756 54 Uppsala (SE). STYMNE, Sten [SE/SE]; Torriösa 1380, S-269 90 Svalöv (SE).
- (74) Agent: FITZNER, Uwe; Lintorfer Str. 10, D-40878 Ratingen (DE).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

Without international search report and to be republished upon receipt of that report.

(54) Title: A NEW CLASS OF ENZYMES IN THE BIOSYNTHETIC PATHWAY FOR THE PRODUCTION OF TRIACYLGLYCEROL AND RECOMBINANT DNA MOLECULES ENCODING THESE ENZYMES



(57) Abstract

The present invention relates to the isolation, identification and characterization of nucleotide sequences encoding an enzyme catalysing the transfer of fatty acids from phospholipids to diacylglycerol in the biosynthetic pathway for the production of triacylglycerol, to the said enzymes and a process for the production of triacylglycerols.

20

25

 \cdot)

A NEW CLASS OF ENZYMES IN THE BIOSYNTHETIC PATHWAY FOR THE PRODUCTION OF TRIACYLGLYCEROL AND RECOMBINANT DNA MOLECULES ENCODING THESE ENZYMES

- The present invention relates to the isolation, identification and characterization of recombinant DNA molecules encoding enzymes catalysing the transfer of fatty acids from phospholipids to diacylglycerol in the biosynthetic pathway for the production of triacylglycerol.
 - Triacylglycerol (TAG) is the most common lipid-based energy reserve in nature. The main pathway for synthesis of TAG is believed to involve three sequential acyl-transfers from acyl-CoA to a glycerol backbone (1, 2). For many years, acyl-CoA: diacylglycerol acyltransferase (DAGAT), which catalyzes the third acyl transfer reaction, was thought to be the only unique enzyme involved in TAG synthesis. It acts by diverting diacylglycerol (DAG) from membrane lipid synthesis into TAG (2). Genes encoding this enzyme were recently identified both in the mouse (3) and in plants (4, 5), and the encoded proteins were shown to be homologous to acyl-CoA: cholesterol acyltransferase (ACAT). It was also recently reported that another DAGAT exists in the oleaginous fungus Mortierella ramanniana, which is unrelated to the mouse DAGAT, the ACAT gene family or to any other known gene (6).

The instant invention relates to novel type of enzymes and their encoding genes for transformation. More specifically, the invention relates to use of a type of genes encoding a not previously described type of enzymes hereinafter designated phospholipid:diacylglycerol acyltransferases (PDAT), whereby this enzyme catalyses an acyl-CoA-independent reaction. The said type of genes expressed alone in transgenic organisms will enhance the total amount of oil (triacylglycerols) produced in the cells. The PDAT genes, in combination with a gene for the synthesis of an uncommon fatty acid will, when expressed in transgenic organisms, enhance the levels of the uncommon fatty acids in the triacylglycerols.

There is considerable interest world-wide in producing chemical feedstock, such as fatty acids, for industrial use from renewable plant resources rather than non-renewable petrochemicals. This concept has broad appeal to manufacturers and consumers on the basis of resource conservation and provides significant opportunity to develop new industrial crops for agriculture.

There is a diverse array of unusual fatty acids in oils from wild plant species and these have been well characterised. Many of these acids have industrial potential and this has led to interest in domesticating relevant plant species to enable agricultural production of particular fatty acids.

Development in genetic engineering technologies combined with greater understanding of the biosynthesis of unusual fatty acids now makes it possible to transfer genes coding for key enzymes involved in the synthesis of a particular fatty acid from a wild species into domesticated oilseed crops. In this way individual fatty acids can be produced in high purity and quantities at moderate costs.

In all crops like rape, sunflower, oilpalm etc., the oil (i.e. triacylglycerols) is the most valuable product of the seeds or fruits and other compounds like starch, protein, and fibre is regarded as by-products with less value. Enhancing the quantity of oil per weight basis at the expense of other compounds in oil crops would therefore increase the value of crop. If genes, regulating the allocation of reduced carbon into the production of oil can be up-regulated, the cells will accumulate more oil on the expense of other products. Such genes might not only be used in already high oil producing cells, such as oil crops, but could also induce significant oil production in moderate or low oil containing crops such as e.g. soy, oat, maize, potato, sugarbeats, and turnips as well as in micro-organisms.

10

15

Summary of the invention

Many of the unusual fatty acids of interest, e.g. medium chain fatty acids, hydroxy fatty acids, epoxy fatty acids and acetylenic fatty acids, have physical properties that are distinctly different from the common plant fatty acids. The present inventors have found that, in plant species naturally accumulating these uncommon fatty acids in their seed oil (i.e. triacylglycerol), these acids are absent, or present in very low amounts in the membrane (phospho)lipids of the seed. The low concentration of these acids in the membrane lipids is most likely a prerequisite for proper membrane function and thereby for proper cell functions. One aspect of the invention is that seeds of transgenic crops can be made to accumulate high amounts of uncommon fatty acids if these fatty acids are efficiently removed from the membrane lipids and channelled into seed triacylglycerols.

PCT/EP00/02701

15

25

30

10

The inventors have identified a novel class of enzymes in plants catalysing the transfer of fatty acids from phospholipids to diacylglycerol in the production of triacylglycerol through an acyl-CoA-independent reaction and that these enzymes (phospholipid:diacylglycerol acyltransferases, abbreviated as PDAT) are involved in the removal of hydroxylated, epoxygenated fatty acids, and probably also other uncommon fatty acids such as medium chain fatty acids, from phospholipids in plants.

This enzyme reaction was shown to be present in microsomal preparations from baker's yeast (*Saccharomyces cerevisiae*). The instant invention further pertains to an enzyme comprising an amino acid sequence as set forth in SEQ ID No. 2 or a functional fragment, derivate, allele, homolog or isoenzyme thereof. A so called ,knock out' yeast mutant, disrupted in the respective gene was obtained and microsomal membranes from the mutant was shown to totally lack PDAT activity. Thus, it was proved that the disrupted gene encodes a PDAT enzyme (SEQ ID NO. 1 and 2). Furtherm, this PDAT enzyme is

10.

15

20

25

30

characterized through the amino acid sequence as set forth in SEQ ID NO 2 containing a lipase motif of the conserved sequence string FXKWVEA.

The instant invention pertains further to an enzyme comprising an amino acid sequence as set forth in SEQ ID NO. 1a, 2b or 5a or a functional fragment, derivate, allele, homolog or isoenzyme thereof.

Further genes and/or proteins of so far unknown function were identified and are contemplated within the scope of the instant invention. A gene from Schizosaccharomyces pombe, SPBC776.14 (SEQ ID. NO. 3), a putative open reading frame CAA22887 of the SPBC776.14 (SEQ ID NO. 13) were identified.

Further Arabidopsis thaliana genomic sequences (SEQ ID NO. 4, 10 and 11) coding for putative proteins were identified, as well as a putative open reading frame AAC80628 from the A. thaliana locus AC 004557 (SEQ ID NO. 14) and a putative open reading frame AAD10668 from the A. thaliana locus AC 003027 (SEQ ID NO. 15) were identified.

Also, a partially sequenced cDNA clone from Neurospora crassa (SEQ ID NO. 9) and a Zea mays EST (Extended Sequence Tac) clone (SEQ ID NO. 7) and corresponding putative amino acid sequence (SEQ ID NO. 8) were identified. Finally, two cDNA clones were identified, one Arabidopsis thaliana EST (SEQ ID NO. 5 and corresponding predicted amino acid sequence SEQ ID NO. 6) and a Lycopersicon esculentum EST clone (SEQ ID NO. 12) were identified. Further, enzymes designated as PDAT comprising an amino acid sequence selected from the group consisting of sequences as set forth in SEQ ID NO 2a, 3a, 5b, 6 or 7b containing a lipase motif FXKWVEA are contemplated within the scope of the invention. Moreover, an enzyme comprising an amino acid sequence encoded through a nucleotide sequence, a portion, derivate, allele or homolog thereof selected from the group consisting of sequences as set forth in SEQ ID No. 1, 1b, 3, 3b, 4, 4a, 4b, 5, 5b, 6b, 7, 8b, 9, 9b, 10, 10b, 11, 11b or 12 or a functional fragment, derivate, allele, homolog or isoenzyme of the enzyme encoding amino acid sequence are included within the scope of the invention.

WO 00/60095 5 PCT/EP00/02701

A functional fragment of the instant enzyme is understood to be any polypeptide sequence which shows specific enzyme activity of a phospholipid:diacylglycerol acyltransferase (PDAT). The length of the functional fragment can for example vary in a range from about 660 ± 10 amino acids to 660 ± 250 amino acids, preferably from about 660 ± 50 to 660 ± 100 amino acids, whereby the "basic number" of 660 amino acids corresponds in this case to the polypeptide chain of the PDAT enzyme of SEQ ID NO. 2 encoded by a nucleotide sequence according to SEQ ID NO. 1. Consequently, the "basic number" of functional fullength enzyme can vary in correspondance to the encoding nucleotide sequence.

A portion of the instant nucleotide sequence is meant to be any nucleotide sequence encoding a polypeptid which shows specific activity of a phospholipid:diacylglycerol acyltransferase (PDAT). The length of the nucleotide portion can vary in a wide range of about several hundreds of nucleotides based upon the coding region of the gene or a highly conserved sequence. For example the length varies in a range form about 1900 ± 10 to 1900 ± 1000 nucleotides, preferably form about 1900 ± 50 to 1900 ± 700 and more preferably form about 1900 ± 100 to 1900 ± 500 nucleotides, whereby the "basic number" of 1900 nucleotides corresponds in this case to the encoding nucleotide sequence of the PDAT enzyme of SEQ ID NO. 1. Consequently, the "basic number" of functional fullength gene can vary.

An allelic variant of the instant nucleotide sequence is understood to be any different nucleotide sequence which encodes a polypeptide with a functionally equivalent function. The alleles pertain naturally occuring variants of the instant nucleotide sequences as well as synthetic nucleotide sequences produced by methods known in the art. Contemplated are even altered nucleotide sequences which result in an enzyme with altered activity and/or regulation or which is resistant against specific inhibitors. The instant invention further includes natural or synthetic mutations of the originally isolated nucleotide

10

15

20

sequences. These mutations can be substitution, addition, deletion, inversion or insertion of one or more nucleotides.

A homologous nucleotide sequence is understood to be a complementary sequence and/or a sequence which specifically hybridizes with the instant nucleotide sequence. Hybridizing sequences include similar sequences selected from the group of DNA or RNA which specifically interact to the instant nucleotide sequences under at least moderate stringency conditions which are known in the art. A preferred, non-limiting example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65°C. This further includes short nucleotide sequences of e.g. 10 to 30 nucleotides, preferably 12 to 15 nucleotides. Included are also primer or hybridization probes.

15

25

30

10

A homologous nucleotide sequence included within the scope of the instant invention is a sequence which is at least about 40%, preferably at least about 50 % or 60%, and more preferably at least about 70%, 80% or 90% and most preferably at least about 95%, 96%, 97%, 98% or 99% or more homologous to a nucleotide sequence of SEQ ID NO. 1.

All of the aforementioned definitions are true for amino acid sequences and functional enzymes and can easily transferred by a person skilled in the art.

Isoenzymes are understood to be enzymes which have the same or a similar substrate specifity and/or catalytic activity but a different primary structure.

In a first embodiment, this invention is directed to nucleic acid sequences that encode a PDAT. This includes sequences that encode biologically active PDATs as well as sequences that are to be used as probes, vectors for transformation or cloning intermediates. The PDAT encoding sequence may

encode a complete or partial sequence depending upon the intended use. All or a portion of the genomic sequence, cDNA sequence, precursor PDAT or mature PDAT is intended.

Further included is a nucleotide sequence selected from the group consisting of sequences set forth in SEQ ID No. 1, 1b, 3, 3b, 4, 4a, 4b, 9b, 10, 10b or 11 or a portion, derivate, allele or homolog thereof. The invention pertains a partial nucleotide sequence corresponding to a fullength nucleotide sequence selected from the group consisting of sequences set forth in SEQ ID No. 5, 5b, 6b, 7, 8b, 9, 11b or 12 or a portion, derivate, allele or homolog thereof. Moreover, a nucleotide sequence comprising a nucleotide sequence which is at least 40% homologous to a nucleotide sequence selected form the group consisting of those sequences set forth in SEQ ID No. 1 1b, 3, 3b, 4, 4a, 4b, 5, 5b, 6b, 7, 8b, 9, 9b, 10, 10b, 11, 11b or 12 is contemplated within the scope of the invention.

15

20

25

30 ·

10

The instant invention pertains to a gene construct comprising a said nucleotide sequences of the instant invention which is operably linked to a heterologous nucleic acid.

The term operably linked means a serial organisation e.g. of a promotor, coding sequence, terminator and/or further regulatory elements whereby each element can fulfill its original function during expression of the nucleotide sequence.

Further, a vector comprising of a said nucleotide sequence of the instant invention is contemplated in the instant invention. This includes also an expression vector as well as a vector further comprising a selectable marker gene and/or nucleotide sequences for the replication in a host cell and/or the integration into the genome of the host cell.

In a different aspect, this invention relates to a method for producing a PDAT in a host cell or progeny thereof, including genetically engineered oil seeds, yeast and moulds or any other oil accumulating organism, via the expression of a

20

construct in the cell. Cells containing a PDAT as a result of the production of the PDAT encoding sequence are also contemplated within the scope of the invention.

Further, the invention pertains a transgenic cell or organism containing a said nucleotide sequence and/or a said gene construct and/or a said vector. The object of the instant invention is further a transgenic cell or organism which is an eucaryotic cell or organism. Preferably, the transgenic cell or organism is a yeast cell or a plant cell or a plant. The instant invention further pertains said transgenic cell or organism having an altered biosynthetic pathway for the production of triacylglycerol. A transgenic cell or organism having an altered oil content is also contemplated within the scope of this invention.

Further, the invention pertains a transgenic cell or organism wherein the activity of PDAT is altered in said cell or organism. This altered activity of PDAT is characterized by an alteration in gene expression, catalytic activity and/or regulation of activity of the enzyme. Moreover, a transgenic cell or organism is included in the instant invention, wherein the altered biosynthetic pathway for the production of triacylglycerol is characterized by the prevention of accumulation of undesirable fatty acids in the membrane lipids.

In a different embodiment, this invention also relates to methods of using a DNA sequence encoding a PDAT for increasing the oil-content within a cell.

Another aspect of the invention relates to the accommodation of high amounts of uncomman fatty acids in the triacylglycerol produced within a cell, by introducing a DNA sequence producing a PDAT that specifically removes these fatty acids from the membrane lipids of the cell and channel them into triacylglycerol. Plant cells having such a modification are also contemplated herein.

Further, the invention pertains a process for the production of triacylglycerol, comprising growing a said transgenic cell or organism under conditions whereby the said nucleotide sequence is expressed and whereby the said transgenic cells comprising a said enzyme catalysing the transfer of fatty acids from phospholipids to diacylglycerol forming triacylglycerol.

Moreover, triacylglycerols produced by the aforementioned process are included in scope of the instant invention.

Object of the instant invention is further the use of an instant nucleotide sequence and/or a said enzyme for the production of triacylglycerol and/or triacylglycerols with uncommon fatty acids. The use of a said instant nucleotide sequence and/or a said enzyme of the instant invention for the transformation of any cell or organism in order to be expressed in this cell or organism and result in an altered, preferably increased oil content of this cell or organism is also contemplated within the scope of the instant invention.

A PDAT of this invention includes any sequence of amino acids, such as a protein, polypeptide or peptide fragment obtainable from a microorganism, animal or plant source that demonstrates the ability to catalyse the production of triacylglycerol from a phospholipid and diacylglycerol under enzyme reactive conditions. By "enzyme reactive conditions" is meant that any necessary conditions are available in an environment (e.g., such factors as temperature, pH, lack of inhibiting substances) which will permit the enzyme to function.

25

30

10

15

20

Other PDATs are obtainable from the specific sequences provided herein. Furthermore, it will be apparent that one can obtain natural and synthetic PDATs, including modified amino acid sequences and starting materials for synthetic-protein modelling from the examplified PDATs and from PDATs which are obtained through the use of such examplified sequences. Modified amino acid sequences include sequences that have been mutated, truncated,

increased and the like, whether such sequences were partially or wholly synthesised. Sequences that are actually purified from plant preparations or are identical or encode identical proteins thereto, regardless of the method used to obtain the protein or sequence, are equally considered naturally derived.

Further, the nucleic acid probes (DNA and RNA) of the present invention can be used to screen and recover "homologous" or "related" PDATs from a variety of plant and microbial sources.

10

Further, it is also apparent that a person skilled in the art can, with the information provided in this application, in any organism identify a PDAT activity, purify an enzyme with this activity and thereby identify a "non-homologous" nucleic acid sequence encoding such an enzyme.

15

The present invention can be essentially characterized by the following aspects:

- 1. Use of a PDAT gene (genomic clone or cDNA) for transformation.
- 2. Use of a DNA molecule according to item 1 wherein said DNA is used for transformation of any organism in order to be expressed in this organism and result in an active recombinant PDAT enzyme in order to increase oil content of the organism.
- 3. Use of a DNA molecule of item 1 wherein said DNA is used for transformation of any organism in order to prevent the accumulation of undesirable fatty acids in the membrane lipids.
 - 4. Use according to item 1, wherein said PDAT gene is used for transforming transgenic oil accumulating organisms engineered to produce any uncommon fatty acid which is harmful if present in high amounts in membrane lipids, such as medium chain fatty acids, hydroxylated fatty acids, epoxygenated fatty acids and acetylenic fatty acids.

10

15

- 5. Use according to item 1, wherein said PDAT gene is used for transforming organisms, and wherein said organisms are crossed with other oil accumulating organisms engineered to produce any uncommon fatty acid which is harmful if present in high amounts in membrane lipids, comprising medium chain fatty acids, hydroxylated fatty acids, epoxygenated fatty acids and acetylenic fatty acids.
- 6. Use according to item 1, wherein the enzyme encoded by said PDAT gene or cDNA is coding for a PDAT with distinct acyl specificity.
- 7. Use according to item 1 wherein said PDAT encoding gene or cDNA, is derived from Saccharornyces cereviseae, or contain nucleotide sequences coding for an amino acid sequence 30% or more identical to the amino acid sequence of PDAT as presented in SEQ. ID. NO. 2.
- 8. Use according to item 1 wherein said PDAT encoding gene or cDNA is derived from *Saccharornyces cereviseae*, or contain nucleotide sequences coding for an amino acid sequence 40% or more *Identical* to the amino acid sequence of PDAT as presented in SEQ. ID. NO. 2.
- Use according to item 1 wherein said PDAT encoding gene or cDNA is derived from Saccharornyces cereviseae, or contain nucleotide sequences coding for an amino acid sequence 60% or more identical to the amino acid sequence of PDAT as presented in SEQ. ID. NO. 2.
- 10. Use according to item 1 wherein sald PDAT encoding gene or cDNA is derived from Saccharomyces cereviseae, or contain nucleotide sequences coding for an amino acid sequence 80% or more identical to the amino acid sequence of PDAT as presented in SEQ. ID. NO. 2.
- 25 11. Use according to item 1 wherein said PDAT encoding gene or cDNA is derived from plants or contain nucleotide sequences coding for an amino acid sequence 40% or more identical to the amino acid sequence of PDAT from *Arabidopsis thaliana* or to the protein encoded by the fullength counterpart of the partial Zea mays, Lycopericon esculentum, or Neurospora crassa cDNA clones.

10

15

- 12. Transgenic oil accumulating organisms comprising, in their genome, a PDAT gene transferred by recombinant DNA technology or somatic hybridization.
- 13. Transgenic oil accumulating organisms according to item 12 comprising, in their genome, a PDAT gene having specificity for substrates with a particular uncommon fatty acid and the gene for said uncommon fatty acid.
- 14. Transgenic organisms according to item 12 or 13 which are selected from the group consisting of fungi, plants and animals.
- 15. Transgenic organisms according to item 12 or 13 which are selected from the group of agricultural plants.
- 16. Transgenic organisms according to item 12 or 13 which are selected from the group of agricultural plants and where said PDAT gene is expressed under the control of a storage organ specific promotor.
- 17. Transgenic organisms according to item 12 or 13 which are selected from the group of agricultural plants and where said PDAT gene is expressed under the control of a seed promotor.
 - 18. Oils from organisms according to item 12 17.
 - 19. A method for altering acyl specificity of a PDAT by alteration of the nucleotide sequence of a naturally occurring encoding gene and as a consequence of this alternation creating a gene encoding for an enzyme with novel acyl specifity.
 - 20. A protein encoded by a DNA molecule according to item 1 or a functional fragment thereof.
 - 21. A protein of item 20 designated phospholipid:diacylglycerol acyltransferase.
- 22. A protein of item 21 which has a distinct acyl specificity.
 - 23. A protein of item 13 having the amino acid sequence as set forth in SEQ, ID NO. 2, 13, 14 or 15 (and the proteins encoded by the fullength or partial genes set forth in SEQ. ID. NO. 1, 3, 4, 5, 7, 9, 10, 11 or 12) or an amino acid sequence with at least 30 % homology to said amino acid sequence.
- 24. A protein of item 23 isolated from Saccharomyces cereviseae.

13 PCT/EP00/02701 WO 00/60095

General methods:

10

15

20

25

Yeast strains and plasmids. The wild type yeast strains used were either FY1679 (MATα his3-Δ200 leu2-Δ1 trp1-Δ6 ura3-52) or W303-1A (MATa ADE2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 ura3-1) (7). The YNR008w::KanMX2 disruption strain FVKT004-04C(AL), which is congenic to FY1679, was obtained from the Euroscarf collection (8). A 2751 bp fragment containing the YNR008w gene with 583 bp of 5' and 183 bp of 3' flanking DNA was amplified genomic DNA using Tag polymerase from TCTCCATCTTCTGCAAAACCT-3' and 5'-CCTGTCAAAAACCTTCTCCTC-3' as primers. The resulting PCR product was purified by agarose gel electrophoresis and cloned into the EcoRV site of pBluescript (pbluescript-pdat). For complementation experiments, the cloned fragment was released from pBluescript by Hindill-Sacl digestion and then cloned between the Hindill and SacI sites of pFL39 (9), thus generating pUS1. For overexpression of the PDAT gene, a 2202 bp EcoRI fragment from the pBluscript plasmid which contains only 24 bp of 5' flanking DNA was cloned into the BamHI site of the GAL1-TPK2 expression vector pJN92 (12), thus generating pUS4.

Microsomal preparations. Microsomes from developing seeds of sunflower (Helianthus annuus), Ricinus communis and Crepis palaestina were prepared using the procedure of Stobart and Stymne (11). To obtain yeast microsomes, 1g of yeast cells (fresh weight) was re-suspended in 8 ml of ice-cold buffer (20 mM Tris-Cl, pH 7.9, 10 mM MgCl₂, 1 mM EDTA, 5 % (v/v) glycerol, 1 mM DTT, 0.3 M ammonium sulfate) in a 12 ml glass tube. To this tube, 4 ml of glass beads (diameter 0.45-0.5 mm) were added, and the tube was then heavily shaken (3 x 60 s) in an MSK cell homogenizer (B. Braun Melsungen AG, Germany). The homogenized suspension was centrifuged at 20,000 x g for 15 min at 6°C and the resulting supernatant was again centrifuged at 100,000 x g 30 for 2 h at 6°C. The 100,000 x g pellet was resuspended in 0.1 M potassium

phosphate (pH 7.2), and stored at -80°C. It is subsequently referred to as the crude yeast microsomal fraction.

14

Lipid substrates. Radio-labeled ricinoleic (12-hydroxy-9-octadecenoic) and vernolic (12,13-epoxy-9-octadecenoic) acids were synthesized enzymatically from [1-14C]oleic acid and [1-14C]linoleic acid, respectively, by incubation with microsomal preparations from seeds of Ricinus communis and Crepis palaestina, respectively (12). The synthesis of phosphatidylcholines (PC) or phosphatidylethanolamines (PE) with ¹⁴C-labeled acyl groups in the sn-2 position was performed using either enzymatic (13), or synthetic (14) acylation of [14Cloleic, [14Clricinoleic, or [14C]vernolic acid. Dioleoyl-PC that was labeled in the sn-1 position was synthesized from sn-1-[14C]oleoyl-lyso-PC and unlabeled oleic acid as described in (14). Sn-1-oleoyl-sn-2-[14C]ricinoleoyl-DAG was synthesized from PC by the action of phospholipase C type XI from B. Cereus (Sigma Chemical Co.) as described in (15). Monovernoloyl- and divernoleoyl-DAG were synthesized from TAG extracted from seeds of Euphorbia lagascae, using the TAG-lipase (Rizhopus arrhizus, Sigma Chemical Co.) as previously described (16). Monoricinoleoyl-TAG was synthesized according to the same method using TAG extracted from Castor bean.

20

25

10

15

Lipid analysis: Total lipid composition of yeast were determined from cells harvested from a 40 ml liquid culture, broken in a glass-bead shaker and extracted into chloroform as described by Bligh and Dyer (17), and then separated by thin layer chromatography in hexane/diethylether/acetic acid (80:20:1) using pre-coated silica gel 60 plates (Merck). The lipid areas were located by brief exposure to I₂ vapors and identified by means of appropriate standards. Polar lipids, sterol-esters and triacylglycerols, as well as the remaining minor lipid classes, referred to as other lipids, were excised from the plates. Fatty acid methylesters were prepared by heating the dry excised material at 85 °C for 60 min in 2% (v/v) sulfuric acid in dry methanol. The methyl esters were extracted with hexane and analyzed by GLC through a 50 m

x 0.32 mm CP-Wax58-CB fused-silica column (Chrompack), with methylheptadecanoic acid as an internal standard. The fatty acid content of each fraction was quantified and used to calculate the relative amount of each lipid class. In order to determine the total lipid content, 3 ml aliquots from yeast cultures were harvested by centrifugation and the resulting pellets were washed with distilled water and lyophilized. The weight of the dried cells was determined and the fatty acid content was quantified by GLC-analyses after conversion to methylesters as described above. The lipid content was then calculated as nmol fatty acid (FA) per mg dry weight yeast.

10

15

20

25

30

Enzyme assays. Aliquots of crude microsomal fractions (corresponding to 10 nmol of microsomal PC) from developing plant seeds or yeast cells were lyophilized over night. ¹⁴C-Labeled substrate lipids dissolved in benzene were then added to the dried microsomes. The benzene was evaporated under a stream of N₂, leaving the lipids in direct contact with the membranes, and 0.1 ml of 50 mM potassium phosphate (pH 7.2) was added. The suspension was thoroughly mixed and incubated at 30°C for the time period indicated, up to 90 min. Lipids were extracted from the reaction mixture using chloroform and separated by thin layer chromatography in hexane/diethylether/acetic acid (35:70:1.5) using silica gel 60 plates (Merck). The radioactive lipids were visualized and quantified on the plates by electronic autoradiography (Instant Imager, Packard, US).

<u>Yeast cultivation</u>. Yeast cells were grown at 28°C on a rotatory shaker in liquid YPD medium (1% yeast extract, 2% peptone, 2% glucose), synthetic medium (18) containing 2% (v/v) glycerol and 2% (v/v) ethanol, or minimal medium (19) containing 16 g/l of glycerol.

The instant invention is further characterized by the following examples which are not limiting:

15

.20

25

30

Acyl-CoA-independent synthesis of TAG by oil seed microsomes. A large number of unusual fatty acids can be found in oil seeds (20). Many of these fatty acids, such as ricinoleic (21) and vernolic acids (22), are synthesized using phosphatidylcholin (PC) with oleoyl or linoleoyl groups esterified to the sn-2 position, respectively, as the immediate precursor. However, even though PC can be a substrate for unusual fatty acid synthesis and is the major membrane lipids in seeds, unusual fatty acids are rarely found in the membranes. Instead, they are mainly incorporated into the TAG. A mechanism for efficient and selective transfer of these unusual acyl groups from PC into TAG must therefore exist in oil seeds that accumulate such unusual fatty acids. This transfer reaction was biochemically characterized in seeds from castor bean (Ricinus communis) and Crepis palaestina, plants which accumulate high levels of ricinoleic and vernolic acid, respectively, and sunflower (Helianthus annuus), a plant which has only common fatty acids in its seed oil. Crude microsomal fractions from developing seeds were incubated with PC having ¹⁴C-labeled oleoyl, ricinoleoyl or vernoloyl groups at the sn-2 position. After the incubation, lipids were extracted and analyzed by thin layer chromatography. We found that the amount of radioactivity that was incorporated into the neutral lipid fraction increased linearly over a period of 4 hours (data not shown). The distribution of [14C]acyl groups within the neutral lipid fraction was analyzed after 80 min (Fig. 1). Interestingly the amount and distribution of radioactivity between diffferent neutral lipids were strongly dependent both on the plant species and on the type of [14C]acyl chain. Thus, sunflower microsomes incorporated most of the label into DAG, regardless of the type of [14Clacvl group. In contrast, R. communis microsomes preferentially incorporated [14C]ricinoleoyl and [14C]vernoloyl groups into TAG, while [14C]oleyl groups mostly were found in DAG. C. palaestina microsomes, finally, incorporated only [14C]vernolyol groups into TAG, with [14C]ricinoleyl groups being found mostly as free fatty acids, and [14C]oleyl groups in DAG. This shows that the high in vivo levels of ricinoleic acid and vernolic acid in the TAG pool of R. communis

WO 00/60095

10

15

20

25

and *C. palaestina*, respectively, can be explained by an efficient and selective transfer of the corresponding acyl groups from PC to TAG in these organisms.

The in-vitro synthesis of triacylglycerols in microsomal preparations of developing castor bean is summarized in table 1.

<u>PDAT: a novel enzyme that catalyzes acyl-CoA independent synthesis of TAG.</u> It was investigated if DAG could serve both as an acyl donor as well as an acyl acceptor in the reactions catalyzed by the oil seed microsomes. Thererfore, unlabeled divernoloyl-DAG was incubated with either *sn*-1-oleoyl-*sn*-2-[¹⁴C]ricinoleoyl-DAG or *sn*-1-oleoyl-*sn*-2-[¹⁴C]ricinoleoyl-PC in the presence of *R. communis* microsomes. The synthesis of TAG molecules containing both [¹⁴C]ricinoleoyl and vernoloyl groups was 5 fold higher when [¹⁴C]ricinoleoyl-PC served as acyl donor as compared to [¹⁴C]ricinoleoyl-DAG (fig.1B). These data strongly suggests that PC is the immediate acyl donor and DAG the acyl acceptor in the acyl-CoA-independent formation of TAG by oil seed microsomes. Therefore, this reaction is catalyzed by a new enzyme which we call phospholipid: diacylglycerol acyltransferase (PDAT).

<u>PDAT activity in yeast microsomes.</u> Wild type yeast cells were cultivated under conditions where TAG synthesis is induced. Microsomal membranes were prepared from these cells and incubated with *sn*-2-[¹⁴C]-ricinoleoyl-PC and DAG and the ¹⁴C-labeled products formed were analyzed. The PC-derived [¹⁴C]ricinoleoyl groups within the neutral lipid fraction mainly were found in free fatty acids or TAG, and also that the amount of TAG synthesized was dependent on the amount of DAG that was added to the reaction (Fig.2). The *in vitro* synthesis of TAG containing both ricinoleoyl and vernoloyl groups, a TAG species not present *in vivo*, from exogenous added *sn*-2-[¹⁴C]ricinoleoyl-PC and unlabelled vernoloyl-DAG (Fig. 2, lane 3) clearly demonstrates the existence of an acyl-CoA-independent synthesis of TAG involving PC and DAG as

substrates in yeast microsomal membranes. Consequently, TAG synthesis in yeast can be catalyzed by an enzyme similar to the PDAT found in plants.

The PDAT encoding gene in yeast.

5

10

15

20

25

A gene in the yeast genome (YNR008w) is known, but nothing is known about the function of YNR008w, except that the gene is not essential for growth under normal circumstances. Microsomal membranes were prepared from the yeast strain FVKT004-04C(AL) (8) in which this gene with unknown function had been disrupted. PDAT activity in the microsomes were assayed using PC with radiolabelled fatty acids at the sn-2 position. The activity was found to be completely absent in the disruption strain (Fig. 2 lane 4). Significantly, the activity could be partially restored by the presence of YNR008w on the single groups of Moreover, acyl 2 lane 5). (Fig. plasmid pUS1 phosphatidylethanolamine (PE) were efficiently incorporated into TAG by microsomes from the wild type strain whereas no incorporation occured from this substrate in the mutant strain (data not shown). This shows that YNR008w encodes a yeast PDAT which catalyzes the transfer of an acyl group from the sn-2 position of phospholipids to DAG, thus forming TAG. It should be noted that no cholesterol esters were formed from radioactive PC even in incubations with added ergosterols, nor were the amount of radioactive free fatty acids formed from PC affected by disruption of the YNR008w gene (data not shown). This demonstrates that yeast PDAT do not have cholesterol ester synthesising or phospholipase activities.

Increased TAG content in yeast cells that overexpress PDAT. The effect of overexpressing the PDAT-encoding gene was studied by transforming a wild type yeast strain with the pUS4 plasmid in which the gene is expressed from the galactose-induced GAL1:TPK2 promoter. Cells containing the empty expression vector were used as a control. The cells were grown in synthetic glycerol-ethanol medium, and expression of the gene was induced after either 2 hours (early log phase) or 25 hours (stationary phase) by the addition of

15

20

25

30

galactose. The cells were then incubated for another 21 hours, after which they were harvested and assays were performed. We found that overexpression of PDAT had no significant effect on the growth rate as determined by the optical density. However, the total lipid content, measured as µmol fatty acids per ma yeast dry weight, was 47% (log phase) or 29% (stationary phase) higher in the PDAT overexpressing strain than in the control. Furthermore, the polar lipid and sterolester content was unaffected by overexpression of PDAT. Instead, the elevated lipid content in these cells is entirely due to an increased TAG content (Fig. 3A,B). Thus, the amount of TAG was increased by 2-fold in PDAT overexpressing early log phase cells and by 40% in stationary phase cells. It is interesting to note that a significant increase in the TAG content was achieved by overexpressing PDAT even under conditions (i.e. in stationary phase) where DAGAT is induced and thus contributes significantly to TAG synthesis. In vitro PDAT activity assayed in microsomes from the PDAT overexpressing strain was 7-fold higher than in the control strain, a finding which is consistent with the increased levels of TAG that we observed in vivo (Fig. 3C). These results clearly demonstrate the potential use of the PDAT gene in increasing the oil content in transgenic organisms.

19

Substrate specificity of yeast PDAT. The substrate specificity of yeast PDAT was analyzed using microsomes prepared from the PDAT overexpressing strain (see Fig. 4). The rate of TAG synthesis, under conditions given in figure 4 with di-oleoyl-PC as the acyl-donor, was 0.15 nmol per min and mg protein. With both oleoyl groups of PC labeled it was possible, under the given assay conditions, to detect the transfer of 11 pmol/min of [14C]oleoyl chain into TAG and the formation of 15 pmol/min of lyso-PC. In microsomes from the PDAT-deficient strain, no TAG at all and only trace amounts of lyso-PC was detected, strongly suggesting that yeast PDAT catalyses the formation of equimolar amounts of TAG and lyso-PC when supplied with PC and DAG as substrates. The fact that somewhat more lyso-PC than TAG is formed can be

explained by the presence of a phospholipase in yeast microsomes, which produces lyso-PC and unesterified fatty acids from PC.

The specificity of yeast PDAT for different acyl group positions was investigated by incubating the microsomes with di-oleoyl-PC carrying a [14C]acyl group either at the sn-1 position (Fig. 4A bar 2) or the sn-2 position (Fig. 4A bar 3). We found that the major ¹⁴C-labeled product formed in the former case was lyso-PC, and in the latter case TAG. We conclude that yeast PDAT has a specificity for the transfer of acyl groups from the sn-2 position of the phospholipid to DAG, thus forming sn-1-lyso-PC and TAG. Under the given assay conditions, trace amounts of 14C-labelled DAG is formed from the sn-1 by the reversible action of a CDP-choline : choline PC phosphotransferase. This labeled DAG can then be further converted into TAG by the PDAT activity. It is therefore not possible to distinguish whether the minor amounts of labeled TAG that is formed in the presence of di-oleoyl-PC carrying a [14Clacyl group in the sn-1 position, is synthesized directly from the sn-1-labeled PC by a PDAT that also can act on the sn-1 postion, or if it is first converted to sn-1-labeled DAG and then acylated by a PDAT with strict selectivity for the transfer of acyl groups at the sn-2 position of PC. Taken together, this shows that the PDAT encoded by YNR008w catalyses an acvl transfer from the sn-2 position of PC to DAG, thus causing the formation of TAG and lyso-PC.

The substrate specificity of yeast PDAT was further analyzed with respect to the headgroup of the acyl donor, the acyl group transferred and the acyl chains of the acceptor DAG molecule. The two major membrane lipids of *S. cerevisiae* are PC and PE, and as shown in Fig. 4B (bars 1 and 2), dioleoyl-PE is nearly 4-fold more efficient than dioleoyl-PC as acyl donor in the PDAT-catalyzed reaction. Moreover, the rate of acyl transfer is strongly dependent on the type of acyl group that is transferred. Thus, a ricinoleoyl group at the *sn*-2 position of PC is 2.5 times more efficiently transferred into TAG than an oleoyl

10

15

25

WO 00/60095 21 PCT/EP00/02701

group in the same position (Fig. 4B bars 1 and 3). In contrast, yeast PDAT has no preference for the transfer of vernoloyl groups over oleoyl groups (Fig. 4B bars 1 and 4). The acyl chain of the acceptor DAG molecule also affects the efficiency of the reaction. Thus, DAG with a ricinoleoyl or a vernoloyl group is a more efficient acyl acceptor than dioleoyl-DAG (Fig. 4B bars 1, 5 and 6). Taken together, these results clearly show that the efficiency of the PDAT-catalyzed acyl transfer is strongly dependent on the properties of the substrate lipids.

<u>PDAT genes.</u> Nucleotide and amino acid sequences of several PDAT genes are given as SEQ ID No. 1 through 15. Futher provisional and/or partial sequences are given as SEQ ID NO 1a through 5a and 1b through 11b, respectively. One of the Arabidopsis genomic sequences (SEQ ID NO. 4) identified an Arabidopsis EST cDNA clone; T04806. This cDNA clone was fully characterised and the nucleotide sequence is given as SEQ ID NO. 5. Based on the sequence homology of the T04806 cDNA and the Arabidopsis thaliana genomic DNA sequence (SEQ ID NO 4) it is apparent that an additional A is present at position 417 in the cDNA clone (data not shown). Excluding this nucleotide would give the amino acid sequence depicted in SEQ ID NO. 12.

Increased TAG content in seeds of Arabidopsis thaliana that express the yeast PDAT. For the expression of the yeast PDAT gene in Arabidopsis thaliana an EcoRI fragment from the pBluescript-PDAT was cloned together with napin promotor (25) into the vector pGPTV-KAN (26). A plasmid (pGNapPDAT) having the yeast PDAT gene in the correct orientation was identified and transformed into Agrobacterium tumefaciens. These bacteria were used to transform Arabidopsis thaliana columbia (C-24) plants using the root transformation method (27). Plants transformed with an empty vector were used as controls.

First generation seeds (T1) were harvested and germinated on kanamycin containing medium. Second generation seeds (T2) were pooled from individual plants and their fatty acid contents analysed by quantification of their methyl

30

10

15

WO 00/60095 22 PCT/EP00/02701

esthers by gas liquid chromatography after methylation of the seeds with 2% sulphuric acid in methanol at 85 °C for 1,5 hours. Quantification was done with heptadecanoic acid methyl esters as internal standard.

From the transformation with pGNapPDAT one T1 plant (26-14) gave raise to seven T2 plants of which 3 plants yielded seeds with statistically (in a mean difference two-sided test) higher oil content than seeds from T2 plants generated from T1 plant 32-4 transformed with an empty vector (table 2).

10

20

30

References cited in the description:

- 1. Bell, R. M. & Coleman, R. A. (1980) Annu. Rev. Biochem. 49, 459-487.
- 2. Stymne, S. & Stobart, K. (1987) in *The biochemistry of plants: a comprehensive treatsie, Vol. 9*, ed. Stumpf, P. K. (Academic Press, New York), pp. 175-214.
 - 3. Cases, S. et al. (1998) Proc. Natl. Acad. Sci. U S A 95, 13018-13023.
- 4. Hobbs, D. H., Lu, C. & Hills, M. J. (1999) FEBS Lett. 452, 145-9
- 5. Zou, J., Wei, Y., Jako, C., Kumar, A., Selvaraj, G. & Taylor, D. C. (1999) Plant J. 19, 645-653.
- 6. Lardizabal, K., Hawkins, D., Mai, J., & Wagner, N. (1999) Abstract presented at the Biochem. Mol. Plant Fatty Acids Glycerolipids Symposium, South Lake Tahoe, USA.
- 7. Thomas, B. J. & Rothstein, R. (1989) Cell 56, 619-630.
- 15 8. Entian, K.-D. & Kötter, P. (1998) Meth. Microbiol. 26, 431-449.
 - 9. Kern, L., de Montigny, J., Jund, R. & Lacroute, F. (1990) Gene 88, 149-157.
 - 10. Ronne, H., Carlberg, M., Hu, G.-Z. & Nehlin, J. O. (1991) *Mol. Cell. Biol.* 11, 4876-4884.
 - 11. Stobart, K. & Stymne, S. (1990) in *Method in Plant Biochemistry*, vol 4, eds. Harwood, J. L. & Bowyer, J. R. (Academic press, London), pp. 19-46.
 - 12. Bafor, M., Smith, M. A., Jonsson, L., Stobrt, A. K. & Stymne, S. (1991) Biochem. J. 280, 507-514.
 - 13. Banas, A., Johansson, I. & Stymne, S. (1992) Plant Science 84, 137-144.
 - 14. Kanda, P. & Wells, M. A. (1981) J. Lipid. Res. 22, 877-879.
- 25 15. Ståhl, U., Ek, B. & Stymne, S. (1998) Plant Physiol. 117, 197-205.
 - 16. Stobart, K., Mancha, M. & Lenman M. Dahlqvist, A. & Stymne, S. (1997) Planta 203, 58-66.
 - 17. Bligh, E. G. & Dyer, W. J. (1959) Can. J. Biochem. Physiol. 37, 911-917.
 - 18. Sherman, F., Fink, G. R. & Hicks, J. B. (1986) in *Laboratory Course Manual for Methods in Yeast Genentics* (Cold Spring Harbor Laboratory)
 - 19. Meesters, P. A. E. P., Huijberts, G. N. M. and Eggink, G. (1996) Appl. Microbiol. Biotechnol. 45, 575-579.
 - 20. van de Loo, F. J., Fox, B. G. & Sommerville, C. (1993), in Lipid metabolism in plants, ed. Moore, T. S. (CRC Press, Inc.), pp. 91-126.
- 21. van de Loo, F. J., Broun, P., Turner, S. & Sommerville, S. (1995) Proc. Natl.

- Acad. Sci. U S A 95, 6743-6747.
- 22. Lee, M., Lenman, M., Banas, A., Bafor, M., Singh, S., Schweizer, M., Nilsson, R., Liljenberg, C., Dahlqvist, A., Gummeson, P-O., Sjödahl, S., Green, A., and Stymne, S. (1998) *Science* 280, 915-918.
- 23. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D.
 G. (1997) Nucl. Acids Res. 24, 4876-4882.
 - 24. Saitou, N. & Nei, M. (1987) Mol. Biol. Evol. 4, 406-425.
 - 25. Stålberg, K., Ellerström, M., Josefsson, L., & Rask, L. (1993) Plant Mol. Biol. 23, 671
- 26. Becker, D., Kemper, E., Schell, J., Masterson, R. (1992) Plant Mol. Biol. 20, 1195
 - 27.D. Valvekens, M. Van Montagu, and Van Lusbettens (1988) Proc. Natl. Acad. Sci. U.S.A. 85, 5536

Description of Figures

FIG. 1:

Metabolism of 14C-labeled PC into the neutral lipid fraction by plant microsomes. (A) Microsomes from developing seeds of sunflower, R. communis and C. palaestina were incubated for 80 min at 30°C with PC (8 nmol) having oleic acid in its sn-1 position, and either 14C-labeled oleic, ricinoleic or vernolic acid in its sn-2 position. Radioactivity incorporated in TAG (open bars), DAG (solid bars), and unsterified fatty acids (hatched bars) was quantified using thin layer chromatography followed by electronic autoradiography, and is shown as percentage of added labeled substrate. (B) Synthesis in vitro of TAG carrying two vernoloyl and one [14C]ricinoleoyl group by microsomes from R. communis. The substrates added were unlabeled divernoloyI-DAG (5 nmol), together with either sn-1-oleoyI-sn-2-[14C]ricinoleoyI-DAG (0.4 nmol, 7700 dpm/nmol) or sn-1-oleoyl-sn-2-[14C]ricinoleoyl-PC (0.4 nmol, 7700 dpm/nmol). The microsomes were incubated with the substrates for 30 min at 30°C, after which samples were removed for lipid analysis as described in the section "general methods". The data shown are the average of two experiments.

20

25

30

15

10

FIG. 2.

PDAT activity in yeast microsomes, as visualized by autoradiogram of neutral lipid products separated on TLC. Microsomal membranes (10 nmol of PC) from the wild type yeast strain FY1679 (lanes 1-3), a congenic yeast strain (FVKT004-04C(AL)) that is disrupted for YNR008w (lane 4) or the same disruption strain transformed with the plasmid pUS1, containing the YNR008w gene behind its native promotor (lane 5), were assayed for PDAT activity. As substrates, we used 2 nmol sn-1-oleoyl-sn-2-[14C]ricinoleoyl-PC together with either 5 nmol of dioleoyl-DAG (lanes 2, 4 and 5) or rac-oleoyl-vernoleoyl-DAG (lane 3). The enzymatic assay and lipid analysis was performed as described in Materials and Methods. The cells were precultured for 20 h in liquid YPD

medium, harvested and re-suspended in an equal volume of minimal medium (19) containing 16 g/l glycerol. The cells were then grown for an additional 24 h prior to being harvested. Selection for the plasmid was maintained by growing the transformed cells in synthetic medium lacking uracil (18). Abbreviations: 1-OH-TAG, monoricinoleoyl-TAG; 1-OH-1-ep-TAG, monoricinoleoyl-monovernoloyl-TAG; OH-FA, unesterified ricinoleic acid.

Fig. 3.

5

10

15

20

25

30

Lipid content (A.B) and PDAT activity (C) in PDAT overexpressing yeast cells. The PDAT gene in the plasmid pUS4 was overexpressed from the galactoseinduced GAL1-TPK2 promotor in the wild type strain W303-1A (7). Its expression was induced after (A) 2 hours or (B) 25 hours of growth by the addition of 2% final concentration (w/v) of galactose. The cells were then incubated for another 22 hours before being harvested. The amount of lipids of the harvested cells was determined by GLC-analysis of its fatty acid contents and is presented as µmol fatty acids per mg dry weight in either TAG (open bar), polar lipids (hatched bar), sterol esters (solid bar) and other lipids (striped bar). The data shown are the mean values of results with three independent yeast cultures. (C) In vitro synthesis of TAG by microsomes prepared from yeast cells containing either the empty vector (vector) or the PDAT plasmid (+ PDAT). The cells were grown as in Fig. 3A. The substrate lipids dioleoyl-DAG (2.5 nmol) and sn-1-oleoyl-sn-2-[14C]-oleoyl-PC (2 nmol) were added to aliquots of microsomes (10 nmol PC), which were then incubated for 10 min at 28 °C. The amount of label incorporated into TAG was quantified by electronic autoradiography. The results shown are the mean values of two experiments.

FIG. 4.

Substrate specificity of yeast PDAT. The PDAT activity was assayed by incubating aliquots of lyophilized microsomes (10 nmol PC) with substrate lipids at 30°C for 10 min (panel A) or 90 min (panel B). Unlabeled DAG (2.5 nmol) was used as substrates together with different labeled phospholipids, as shown

in the figure. (A) Sn-position specificity of yeast PDAT regarding the acyl donor substrate. Dioleoyl-DAG together with either sn-1-[14C]oleoyl-sn-2-[14C]oleoyl-PC (di-[14C]-PC), sn-1-[14C]oleoyl-sn-2-oleoyl-PC (sn1-[14C]-PC) or sn-1-oleoylsn-2-[14C]oleoyl-PC (sn2-[14C]-PC). (B) Specificity of yeast PDAT regarding phospholipid headgroup and of the acyl composition of the phospholipid as wellas of the diacylglycerol. Dioleoyl-DAG together with either sn-1-oleoyl-sn-2-[14C]oleoyl-PC (oleoyl-PC), sn-1-oleoyl-sn-2-[14C]oleoyl-PE (oleoyl-PE), sn-1-(ricinoleoyl-PC) sn-1-oleovi-sn-2oleovi-sn-2-[14C]ricinoleoyi-PC or [14C]vernoloyI-PC (vernoloyI-PC). In the experiments presented in the 2 bars to the far right, monoricinoleoyl-DAG (ricinoleoyl-DAG or mono-vernoloyl-DAG (vernoloyi-DAG) were used together with sn-1-oleoyi-sn-2-[14C]-oleoyi-PC. The label that was incorporated into TAG (solid bars) and lyso-PC (LPC, open bars) was quantified by electronic autoradiography. The results shown are the mean values of two experiments. The microsomes used were from W303-1A_cells overexpressing the PDAT gene from the GAL1-TPK2 promotor, as described in Fig. 3. The expression was induced at early stationary phase and the cells were harvested after an additional 24 h.

20 TAB.1:

15

In vitro synthesis of triacylglycerols in microsomal preparations of developing castor bean. Aliquots of microsomes (20 nmol PC) were lyophilised and substrate lipids were added in benzene solution: (A) 0.4 nmol [14 C]-DAG (7760 dpm/nmol) and where indicated 1.6 nmol unlabelled DAG; (B) 0.4 nmol [14 C]-DAG (7760 dpm/nmol) and 5 nmol unlabelled di-ricinoleoyl-PC and (C) 0.25 nmol [14 C]-PC (4000 dpm/nmol) and 5 nmol unlabelled DAG. The benzene was evaporated by N₂ and 0.1 ml of 50 mM potassium phosphate was added, thoroughly mixed and incubated at 30 °C for (A) 20 min.; (B) and (C) 30 min.. Assays were terminated by extraction of the lipids in chloroform. The lipids were then separated by thin layer chromatography on silica gel 60 plates

PCT/EP00/02701

(Merck; Darmstadt, Germany) in hexan/diethylether/acetic 35:70:1.5. The radioactive lipids were visualised and the radioactivity quantified on the plate by electronic autoradiography (Instant Imager, Packard, US). Results are presented as mean values of two experiments.

5

10

Radioactivity in different triacylglycerols (TAG) species formed. Abbreviations used: 1-OH-, mono-ricinoleoyl-; 2-OH, di-ricinoleoyl-; 3-OH-, triricinoleoyl; 1-OH-1-ver-, mono-ricinoleoly-monovernoleoyl-; 1-OH-2-ver-, mono-ricinoleoyl-divernoleoyl-. Radiolabelled DAG and PC were prepared enzymatically. The radiolabelled ricinoleoyl group is attached at the sn-2-position of the lipid and unlabelled oleoyl group at the sn-1-position. Unlabelled DAG with vernoleoyl- or ricinoleoyl chains were prepared by the action of TAG lipase (6) on oil of Euphorbia lagascae or Castor bean, respectively. Synthetic di-ricinoleoyl-PC was kindly provided from Metapontum Agribios (Italy).

15

20

TAB.2:

Total fatty acids per mg of T2 seeds pooled from individual *Arabidopsis thaliana* plants transformed with yeast PDAT gene under the control of napin promotor (26-14) or transformed with empty vector (32-4).

* = stastistical difference between control plants and PDAT transformed plants in a mean difference two-sided test at $\alpha = 5$.

Description of the SEQ ID:

SEQ ID NO. 1: Genomic DNA sequence and suggested amino acid sequence of the Saccharomyces cerevisiae PDAT gene, YNR008w, with GenBank accession number Z71623 and Y13139, and with nucleotide ID number 1302481.

SEQ ID NO. 2: The amino acid sequence of the suggested open reading frame YNR008w from Saccharomyces cerevisiae.

SEQ ID NO. 3: Genomic DNA sequence of the Schizosaccharomyces pombe gene SPBC776.14.

SEQ ID NO. 4: Genomic DNA sequence of part of the Arabidopsis thaliana locus with GenBank accession number AB006704.

15

20

SEQ ID NO. 5: Nucleotide sequence of the Arabidopsis thaliana cDNA clone with GenBank accession number T04806, and nucleotide ID number 315966.

SEQ ID NO. 6: Predicted amino acid sequence of the Arabidopsis thaliana cDNA clone with GenBank accession number T04806.

SEQ ID NO. 7: Nucleotide and amino acid sequence of the Zea mays EST clone with GenBank accession number Al491339, and nucleotide ID number 4388167.

25 SEQ ID NO. 8: Predicted amino acid sequence of the Zea mays EST clone with GenBank accession number Al491339, and nucleotide ID number 4388167.

SEQ ID NO. 9: DNA sequence of part of the Neurospora crassa EST clone W07G1, with GenBank accession number Al398644, and nucleotide ID number 4241729.

SEQ ID NO. 10: Genomic DNA sequence of part of the Arabidopsis thaliana locus with GenBank accession number AC004557.

SEQ ID NO. 11: Genomic DNA sequence of part of the Arabidopsis thaliana locus with GenBank accession number AC003027.

SEQ ID NO. 12: DNA sequence of part of the Lycopersicon esculentum cDNA clone with GenBank accession number Al486635.

SEQ ID NO. 13: Amino acid sequence of the Schizosaccharomyces pombe putative open reading frame CAA22887 of the Schizosaccharomyces pombe gene SPBC776.14.

SEQ ID NO. 14: Amino acid sequence of the Arabidopsis thaliana putative open reading frame AAC80628 derived from the Arabidopsis thaliana locus with GenBank accession number AC004557.

SEQ ID NO 15: Amino acid sequence of the Arabidopsis thaliana putative open reading frame AAD10668 derived from the Arabidopsis thaliana locus with GenBank accession number AC003027.

Further provisional and/or partial sequences are defined through the following SEO IDs:

25 SEQ ID NO. 1a: The amino acid sequence of the yeast ORF YNR008w from Saccharomyces cerevisiae.

SEQ ID NO. 2a: Amino acid sequence of the region of the Arabidopsis thaliana genomic sequence (AC004557).

SEQ ID NO. 3a: Amino acid sequence of the region of the Arabidopsis thaliana genomic sequence (AB006704).

SEQ ID NO. 4a: The corresponding genomic DNA sequence and amino acid sequence of the yeast ORF YNROO8w from Saccharomyces cerevisiae.

SEQ ID NO. 5a: The amino acid sequence of the yeast ORF YNROO8w from Saccharomyces cerevisiae derived form the corresponding genomic DNA sequence.

10

SEQ ID NO. 1b: Genomic DNA sequence of the Saccharomyces cerevisiae PDAT gene, YNR008w, genebank nucleotide ID number 1302481, and the suggested YNR008w amino acid sequence.

15

20

SEQ ID NO. 2b: The suggested amino acid sequence of the yeast gene YNR008w from Saccharomyces cerevisiae.

SEQ ID NO. 3b: Genomic DNA sequence of the Schizosaccharomyces pombe gene SPBC776.14.

SEQ ID NO. 4b: Genomic DNA sequence of part of the Arabidopsis thaliana locus with genebank accession number AB006704.

SEQ ID NO. 5b: Nucleotide sequence and the corresponding amino acid sequence of the Arabidopsis thaliana EST-clone with genebank accession number T04806, and ID number 315966.

SEQ ID NO. 6b: Nucleotide and amino acid sequence of the Zea mays cDNA clone with genebank ID number 4388167.

SEQ ID NO. 7b: Amino acid sequence of the Zea mays cDNA clone with genebank ID number 4388167.

SEQ ID NO. 8b: DNA sequence of part of the Neurospora crassa cDNA clone WO7G1, ID number 4241729.

SEQ ID NO. 9b: Genomic DNA sequence of part of the Arabidopsis thaliana locus with genebank accession number AC004557.

10 SEQ ID NO. 10b: Genomic DNA sequence of part of the Arabidopsis thaliana locus with genebank accession number AC003027.

SEQ ID NO. 11b: DNA sequence of part of the Lycopersicon esculentum cDNA clone with genebank accession number Al486635.

15

20

25

Claims

- An enzyme catalysing in an acyl-CoA-independent reaction the transfer of fatty acids from phospholipids to diacylglycerol in the biosynthetic pathway for the production of triacylglycerol.
- 2. An enzyme according to claim 1, comprising an amino acid sequence as set forth in SEQ ID No. 2 or a functional fragment, derivate, allele, homolog or isoenzyme thereof.

10

20

25

- An enzyme according to claims 1 or 2 designated as phospholipid:diacylglycerol acyltransferase (PDAT).
- 4. An enzyme according to claims 1 to 3, comprising an amino acid sequence as set forth in SEQ ID No. 1a, 2b or 5a or a functional fragment, derivate, allele, homolog or isoenzyme thereof.
 - 5. An enzyme according to claims 1 to 4, comprising an amino acid sequence selected from the group consisting of sequences as set forth in SEQ ID No. 2a, 3a, 5b, 6, 7b, 8, 13, 14 or 15 or a functional fragment, derivate, allele, homolog or isoenzyme thereof.
 - 6. An enzyme according to claims 1 to 5, comprising an amino acid sequence encoded through a nucleotide sequence, a portion, derivate, allele or homolog thereof selected from the group consisting of sequences as set forth in SEQ ID No. 1, 1b, 3, 3b, 4, 4a, 4b, 5, 5b, 6b, 7, 8b, 9, 9b, 10, 10b, 11, 11b or 12 or a functional fragment, derivate, allele, homolog or isoenzyme of the enzyme encoding amino acid sequence.
- 7. A nucleotide sequence encoding an enzyme catalysing in an acyl-CoAindependent reaction the transfer of fatty acids from phospholipids to

10

15

20

diacylglycerol in the biosynthetic pathway for the production of triacylglycerol.

- 8. A nucleotide sequence according to claim 7 encoding an enzyme designated as phospholipid:diacylglycerol acyltransferase (PDAT).
 - 9. A nucleotide sequence according to claims 7 or 8, selected from the group consisting of sequences as set forth in SEQ ID No. 1, 1b, 3, 3b, 4, 4a, 4b, 9b, 10, 10b or 11 or a portion, derivate, allele or homolog thereof.

10. A partial nucleotide sequence corresponding to a fullength nucleotide sequence according to claims 7 to 9, selected from the group consisting of sequences as set forth in SEQ ID No. 5, 5b, 6b, 7, 8b, 9, 11b or 12 or a portion, derivate, allele or homolog thereof.

11. A nucleotide sequence according to claims 7 to 10, comprising a nucleotide sequence which is at least 40% homologous to a nucleotide sequence selected form the group consisting of those sequences set forth in SEQ ID No. 1, 1b, 3, 3b, 4, 4a, 4b, 5, 5b, 6b, 7, 8b, 9, 9b, 10, 10b, 11, 11b or 12.

- 12. A gene construct comprising a nucleotide sequence according to claims 7 to 11 operably linked to a heterologous nucleic acid.
- 13. A vector comprising a nucleotide sequence according to claims 7 to 11 or agene construct according to claim 12.
 - 14. A vector according to claim 13, which is an expression vector.
- 15. A vector according to claims 13 or 14, further comprising a selectable marker gene and/or nucleotide sequences for the replication in a host cell or the integration into the genome of the host cell.

16. A transgenic cell or organism containing a nucleotide sequence according to claims 7 to 11 and/or a gene construct according to claim 12 and/or a vector according to claims 13 to 15.

. 5

17. A transgenic cell or organism according to claim 16 which is an eucaryotic cell or organism.

18. A transgenic cell or organism according to claims 16 or 17 which is a yeast cell or a plant cell or a plant.

10

19. A transgenic cell or organism according to claims 16 to 18 having an altered biosynthetic pathway for the production of triacylglycerol.

15

20. A transgenic cell or organism according to claims 16 to 19 having an altered oil content.

21. A transgenic cell or organism according to claims 16 to 20 wherein the

activity of PDAT is altered.

. 20

22. A transgenic cell or organism according to claims 16 to 21 wherein the altered activity of PDAT is characterized by an alteration in gene expression, catalytic activity and/or regulation of activity of the enzyme.

25

23. A transgenic cell or organism according to claims 16 to 22 wherein the altered biosynthetic pathway for the production of triacylglycerol is characterized by the prevention of accumulation of undesirable fatty acids in the membrane lipids.

30

24. A process for the production of triacylglycerol, comprising growing a transgenic cell or organism according to claims 16 to 23 under conditions whereby the said nucleotide sequence according to claims 7 to 11 is expressed.

- 25. Triacylglycerols produced by a process according to claim 24.
- 26. Use of a nucleotide sequence according to claims 7 to 11 and/or an enzyme according to claims 1 to 6 for the production of triacylglycerol and/or triacylglycerols with uncommon fatty acids.
- 27. Use of a nucleotide sequence according to claims 7 to 11 and/or an enzyme according to claims 1 to 6 for the transformation of any cell or organism in order to be expressed in this cell or organism and result in an altered, preferably increased oil content of this cell or organism.

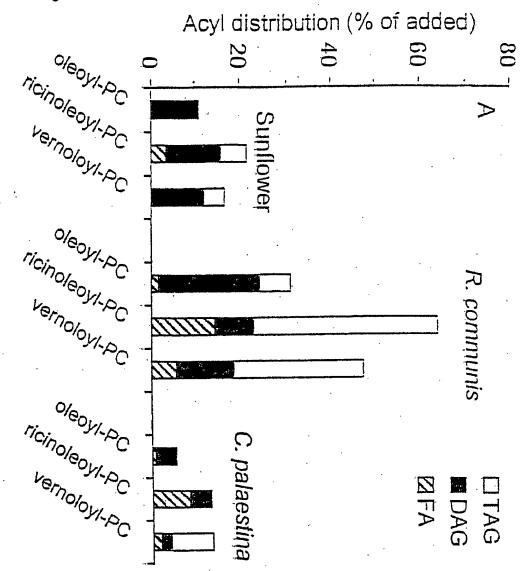
15

5

Figurs

1/6

Fig. 1:



Radioactivity in ricinoleoyl-vernoloyl-TAG (% of added)

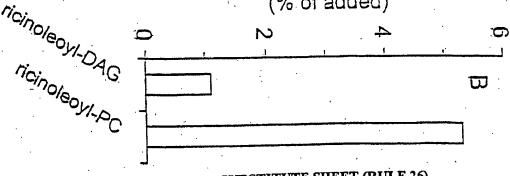
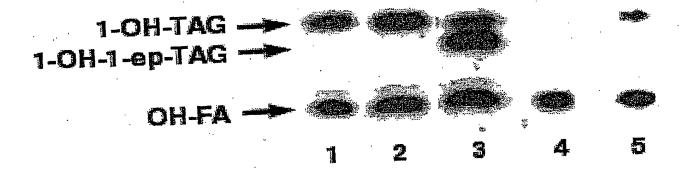
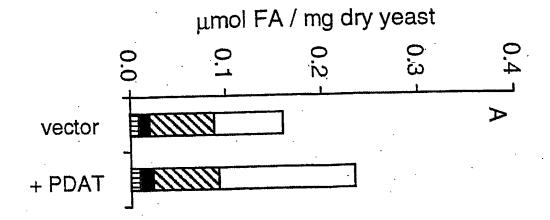
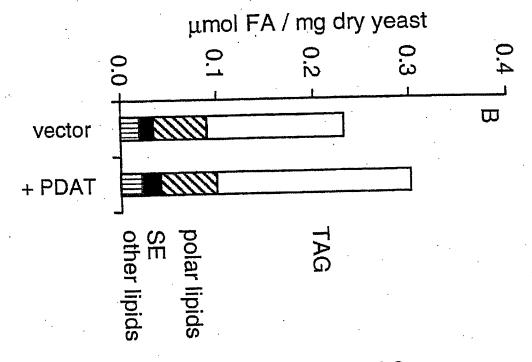
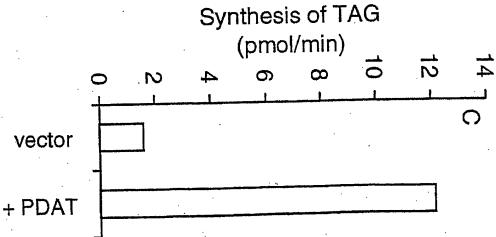


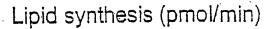
Fig 2

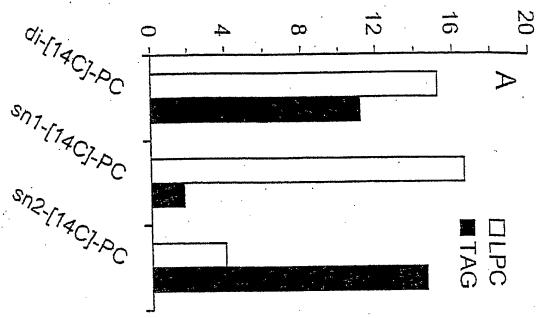




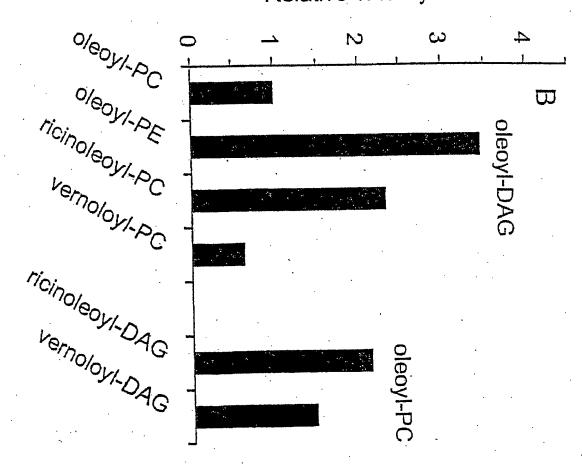








Relative TAG-synthesis



Tab. 2:

T1 plant deviation	T2 plant number	nmol fatty acids per mg seed	standard
32-4	1	1277	±11 (n=2)
	4	1261	±63 (n=3)
	5	1369	$\pm 17 \text{ (n=3)}$
	6	1312	<u>+</u> 53 (n=4)
	7	1197	±54 (n=5)
	8	1240	\pm 78 (n=4)
•	9	1283	$\pm 54 (n=5)$
•	. 10	1381	<u>+</u> 35 (n=5)
26-14	1	1444	<u>+</u> 110 (n=4)
20 1.	$\overline{2}$	1617*	±109 (n=4)
	3	1374	$\pm 37 \text{ (n=2)}$
	5	1562*	$\pm 70 \text{ (n=4)}$
	6	1393	<u>+</u> 77 (n=4)
	7	1433	±98 (n=4)
	8	1581*	±82 (n=4)

Sequence Listing

<210 <211 <212 <213	> 198 > ge	nomi			cere	visi	ae									
	> CD > (1		1983)			·					. •				
<400	> 1															
atg	ggc	aca	ctg	ttt	cga	aga	aat	gtc.	cag	aac	caa	aag	agt	gat	tct	48
Met	Gly	Thr	Leu	Phe	Arg	Arg	Asn	Val	Gln	Asn	Gln	Lys	Ser		Ser	
1				5					10					15		
						ggt	tet.	crt t	cat	aac	aag	cas	gag	aσc	aga	96
gat	Glu	Asn	Asn	Lvs	Glv	Gly	Ser	Val	His	Asn	Lys	Arg	Glu	Ser	Arg	
vab	Gru	ron	20		0-7	,		25			•	_	30			
																1 4 4
						cag										144
Asn	His		His	His	Gln	Gln	G19 40	ren	GIA	HIS	гу	45	ALU	Arg	GīŽ	
		35					40					Ŧ -				
att	agt	ggc	agt	gca	aaa	aga	aat	gag	cgt	ggc	aaa	gat	ttc	gac	agg	192
Ile	Ser	Gly	Ser	Ala	Lys	Arg	Asn	Glu	Arg	Gly	Lys	Asp	Phe	Asp	Arg	
	50					. 55	•				. 60			•		
																240
						aga										
		Ąsp	GIY	ASN	. GIY 70		nλa	wrd	ııp	75		261	A	9	Leu 80	•
65 ^					, 0	•			•							
att	ttc	att	ctt	ggt	gca	ttc	tta	ggt	gta	ctt	ttg	ccg	ttt	ago	ttt	288
Ile	Phe	Ile	Leu	Gly	Ala	Phe	Leu	Gly	Val	Leu	Leu	Pro	Phe	Ser	Phe	
				85	i				90					95	i	
			٠.					. ~-+	3.00				a a c	. aac	. +++	336
ggc	gct	tat	cat oic	gtt	. cat	. aat	. agc	· Asr	Ser	. gac	Len	Phe	Ast	Asr	ttt Phe	
GTA	AId	TĀT	100		. nis	, Ren	Jer	105					110			
gta	aat	ttt	gat	tca	a ctt	aaa	gtg	tat	ttg	gat	gat	tgg	r aaa	a gat	gtt:	384
Va]	. Asn	Phe	Asp	Ser	Lei	ı Lys	val	Tyr	Leu	Ası) Asr	Tr	Lys	Asp	val	
		115	j.			•	120					125	5			
									· ~=+	י מי	- art	. רשר	י מכו	gat	aac	432
cto	. cca	caa	ggt	. ata	a agt	. CC9	. CCC - Phe	Tle	. ye.	. yai Ast	. al.	. caç e Glr	, gov	a Gly	Asn	
net	130		. GTĀ		. nei	135		:	F	<u>-</u>	140			•		

													·			
tac	tcc	aca	tct	tct	tta	gat	gat	ctc	agt	gaa	aa't	ttt	gcc.	gtt	ggt	480
Tyr	Ser	Thr	Ser	Ser	Leu	Asp	Asp	Leu	Ser	Glu	Asn	Phe	Ala	Val	Gly	•
145					150	-	-			155					160	
747					130											
					•						_					E20
						tat										528
Lys	Gln	Leu	Leu	Arg	Asp	Tyr	Asn	Ile	Glu	Ala	Lys	His	Pro	Val	Val	
				165					170					175		
atr	arr	cct	aat	αtc	att	tct	аса	aga	att	σаа	agc	taa	σσа	att	att	576
						Ser										
wet	VZI	PIO		vai	TTE	Set	IIII		TIG	Gin	267	עבי		742	116	
			180			•		185					190			
gga	gac	gat	gag	tgc	gat	agt	tct	gcg	cat	ttt	cgt	aaa	cgg	ctg	tgg	624
Glv	Asp	aeA	Glu	Cys	Asp	Ser	Ser	Ala	His	Phe	Arg	Lys	Arg	Leu	Trp	
•	•	195		-	-		200					205				
		ر. ر ند														
					·					- m						672
						aga										0/4
Gly	Ser	Phe	Tyr	Met	Leu	Arg	Thr	Met	Val	Met			vaı	Cys	Trp	*
	210					215					220					
tta	aaa	cat	σta	aco	tta	gat	cct	gaa	aca	ggt	ctg	gac	cca	cag	aac	720
_															Asn	
		nra	Val	. Mec			110	010	****	235					240	
225					230	•				233	,				240	**
															atc	768
Phe	Thr	Leu	. Arg	, Ala	Ala	Gln	Gly	Phe	Glu	Ser	Thr	Ası	туг	Phe	e Ile	
				245	i				250					255	5	
						•										
		, rat	· + ~	, att	+	1 aac		att	ttc	· caa	a aat	eto	r aas	a ota	a att	816
_																
Ala	GIY	r Tyr	,		rr) ASI	LLYS			9 611	1 ASI	ı ne			l Ile	•
			260)			•	265	•		•	•	270) .		
ggo	: tat	gaa	2 CCC	aat	: aaa	alato	aco	g agt	gct	ge	tat	gat	: tgg	g agg	g ctt	864
G1v	TVI	Gli	ı Pro	Asr	ı Lys	s Met	Thr	Ser	Ala	a Ala	а Туг	: As	o Tri	Arg	g Leu	
_	-	275			_		280					289				
								-					11.			
																013
															a aag	912
Ala	туз	r Lei	ı Ası) Lev	1 G1	ı Arç	Arq	, Ast	Arg	ı Tyı	r Phe	€ Th	r Ly:	s Le	u Lys	
	290)				295	5			•	300)				
													. •			
gaa	a caa	a atr	caa	a chr	r tti	t car	caa	a tto	a a a a	. aai	c gaa	a aa	a gti	t ta	t tta	960
								•							s Leu	
		. <u>.</u>	الدى د	. HE			, 411	, 2000							320	
305)				31	U				31!					320	
att	gga	a cal	tct	tate	g gg	t tct	cag	g att	ato	c tti	t ta	tt	t at	g aa	a tgg	1008
Ilé	e G1y	y His	s Sei	r Met	: Gl	y Sei	Gl:	ı Ile	lle	e Phe	e Ty:	r Ph	e Me	Ly:	s Trp	
	-			325					330		•			33		
					٠.					-						

	•							31.	,,							
							tac Tyr									1056
							att Ile 360	Asn								1104
							cta Leu						-			1152
	•														tca Ser 400	1200
										Trp					tca Ser	1248
				Gly					Trp					Ser	tct Ser	1296
			Ala					Thr					Asn		att lle	1344
		Glu					Asp					Asr			a atg	1392
	Asp	-				Thr					Pro				c caa ı Gln 480	1440
					ı Glr					Ty					a gaa u Glu 5	1488
	_	_		s Ası		-		•	s Lys					ı Pro	a atg o Met	1536
) Le					Hi:					c Cys		a tac e Tyr	1584

														••		1677
ggg	gtg	aac	aac	cca	act	gaa	agg	gca	tat	gta	tat	aag	gaa	gag	gat	1632`
Glv	Val	Asn	Asn	Pro	Thr	Glu	Arg	Ala	Tyr	Val	Tyr	Lys	Glu	Glu	qzA	
1	530					535					540					
	220					J J J										
									~=~	T 2.C	~==	200	224	caa	cct	1680
gac	tcc	tct	gct	ctg	aat	ttg	acc	all	gac	m	944	age	Tue	C3=	Dro	2000
Asp	Ser	Ser	Ala	Leu	Asn	Leu	Thr	ITE	Asp		GIU	Ser	TÃ2	GIII	710	
545					550					555					560	
ata	++0	CEC	acc	σασ	aga	σac	aga	acc	gtt	ccg	ctc	gtg	gcg	cat	tca	1728
y.a	nh-		mh-	C1:	233 C1v	Aen	Glv	Thr	Val	Pro	Leu	Val	Ala	His	Ser	
vaı	Pue	ren	1,717		Gry	Hop	0-1		570					575		
				565					370					3.5		
•																1776
atg	tgt	cac	aaa	. tgg	gcc	cag	ggt	gct	tca	ccg	tac	aac	CCT	gcc	gga	1776
Met	Cvs	His	Lys	Trp	Ala	Gln	Gly	Ala	Ser	Pro	Туг	Asn	Pro	Ala	Gly	
	•		580				•	585					590			
						. ~~~	a + α		cac	cao	r dda	gat	: cas	. ttt	gat	1824
att	aac	gtt	act	. act	. grg	yaa	. alg	Tire	uic	215	D~	Aer	. Arc	r Phe	Asp	
Ile	e Ası	ı Val	Th:	: Ile	Ya1	. GIU			urs	نندي	r Fr	. ASE	Nare	,	g Asp	
		595	5				600	1				605)			
ata	a cat	z aat	c aga	a gca	a aaa	ago	gco	gaa	cac	gta	a ga	c ato	ct(gg	c agc	1872
Tl	- Δr	- 23 - 21s	z G1v	- z Ala	a Lvs	s Ser	Ala	ı Glu	. His	Val	L As	o Il	e Le	ı Gly	y Ser	
**	61		, 01.	,		615					62	0				•
	ρ.T.	U				Ų 1.	•									
									4			~ ~~		- ~~	c dat	1920
gc	g ga	g·tt	g aa	c ga	t ta	c ato	פודני	y aaa	aatt	; gca	a ay	c gg	L aa	- 03	c gat	1,720
Al	a Gl	u Le	u As	n As	p Ty	r Ile	e Lei	ı Lys	: Ile	Ali	a Se	r GI	y As	n GI	y Asp	•
. 62	5				63	0				63	5				640	
		г к э	a cc	a ~~	כ כפ	a tt	a ta	t aai	t tto	a a o	с са	g tg	g gt	t tc	t cag.	1968
	.c gc	c ya	y cc	a cy	~ ~1	- Ta	, Ca	~ Δα:	n T.ei	1 Se	r Gl	n Tr	n Va	1 Se	r Gln	
L€	u va	ı Gl	u Pr			11 TIE	u 56.	_ A-0.						65	ร์	
•	٠.			64	5		•		65	J				0.5	· - .	
																4005
at	g co	c tt	.c cc	a at	g ta	.a										1986
Me	et Pr	o Ph	ie Pr	o Me	t											

```
<210> 2
<211> 661
<212> PRT
<213> Saccharomyces cerevisiae
<400> 2
Met Gly Thr Leu Phe Arg Arg Asn Val Gln Asn Gln Lys Ser Asp Ser
Asp Glu Asn Asn Lys Gly Gly Ser Val His Asn Lys Arg Glu Ser Arg
Asn His Ile His His Gln Gln Gly Leu Gly His Lys Arg Arg Arg Gly
                             40
Ile Ser Gly Ser Ala Lys Arg Asn Glu Arg Gly Lys Asp Phe Asp Arg
                        - 55
Lys Arg Asp Gly Asn Gly Arg Lys Arg Trp Arg Asp Ser Arg Arg Leu
                                          75
                     70
Ile Phe Ile Leu Gly Ala Phe Leu Gly Val Leu Leu Pro Phe Ser Phe
Gly Ala Tyr His Val His Asn Ser Asp Ser Asp Leu Phe Asp Asn Phe
                                 105
Val Asn Phe Asp Ser Leu Lys Val Tyr Leu Asp Asp Trp Lys Asp Val
                             120
        115
Leu Pro Gln Gly Ile Ser Ser Phe Ile Asp Asp Ile Gln Ala Gly Asn
                                             140
                         135
Tyr Ser Thr Ser Ser Leu Asp Asp Leu Ser Glu Asn Phe Ala Val Gly
                     150
                                         155
 Lys Gln Leu Leu Arg Asp Tyr Asn Ile Glu Ala Lys His Pro Val Val
                                     170
                 165
 Met Val Pro Gly Val Ile Ser Thr Gly Ile Glu Ser Trp Gly Val Ile
             180
                                 185
 Gly Asp Asp Glu Cys Asp Ser Ser Ala His Phe Arg Lys Arg Leu Trp
                             200
 Gly Ser Phe Tyr Met Leu Arg Thr Met Val Met Asp Lys Val Cys Trp
                         215
 Leu Lys His Val Met Leu Asp Pro Glu Thr Gly Leu Asp Pro Pro Asn
                     230
                                          235
 Phe Thr Leu Arg Ala Ala Gln Gly Phe Glu Ser Thr Asp Tyr Phe Ile
                                     250
 Ala Gly Tyr Trp Ile Trp Asn Lys Val Phe Gln Asn Leu Gly Val Ile
                                 265
             260
 Gly Tyr Glu Pro Asn Lys Met Thr Ser Ala Ala Tyr Asp Trp Arg Leu
                            .280
         275
 Ala Tyr Leu Asp Leu Glu Arg Arg Asp Arg Tyr Phe Thr Lys Leu Lys
                         295
                                              300
 Glu Gln Ile Glu Leu Phe His Gln Leu Ser Gly Glu Lys Val Cys Leu
                                          315
                     310
 Ile Gly His Ser Met Gly Ser Gln Ile Ile Phe Tyr Phe Met Lys Trp
                 325
                                      330
```

		•														
	Val	Glu	Ala	Glu 340	Gly	Pro	Leu		Gly 345	Asn	Gly	Gly	Arg	Gly 350	Trp	Val
	Asn	Glu	His 355	Ile	Asp	Ser	Phe	Ile 360	Asn	Ala		Gly	Thr 365	Leu	Leu	Gly
	Ala	Pro 370	Lys	Alá	Val	Pro	Ala 375	Leu	Ile	Ser	Gly	Glu 380	Met	Lys	Asp	Thr
	11e 385		Leu	Asn	Thr	Leu 390	Ala	Met	Tyr	Gly	Leu 395	Ģlu	Lys	Phe	Phe	Ser 400
		Ile	Glu	Arg	Val 405		Met	Leu	Gln	Thr 410	Trp	Gly	Gly	Ile	Pro 415	Ser
	Met	Leu	Pro	Lys 420	Gly	Glu	Glu	Val	Ile 425		Gly	Asp	Met	Lys 430	Ser	Ser
	Ser	Glu	Asp	Ala		Asn	Asn	Asn 440	Thr	Asp	Thr	Tyr	Gly 445	Asn	Phe	Ile
	Arg	Phe	Glu	Arg	Asn ·	Thr	Ser 455	Asp	Ala	Phe	Asn	Lys 460	Asn	Leu	Thr	Met
	Lys 465		Ala	. Ile	Asn	Met 470		Leu	Ser	Ile	Ser 475		Glu	Trp	Leu	Gln 480
			Val	His	Glu 485		Tyr	Ser	Phe	Gly 490		Ser	Lys	Asn	Glu 495	Glu
	Glu	Leu	Arg	Lys 500	Asn		Leu	. His	His 505			Trp	Ser	Asn 510	Pro	Met
	Glu	Val	. Pro		Pro	Glu	Ala	. Pro		Met	Lys	Ile	Tyr 525		Ile	Tyr
	Gly	Val	. Ásı		Pro	Thr	Glu 535		Ala	Tyr	Val	Tyr 540		Glu	Glu	Asp
	Asp 545	Ser		Ala	Leu	Asn 550		. Thr	Ile		Tyr		Ser	Lys	Glr	9ro 560
	Val	. Phe	e Lev	ı Thr	Glu 565		Asr	Gly	Thi	.Val		Leu	Va]	Ala	His 575	Ser
•	Met	Cys	s His			Ala			Ala 585		Pro	Tyr	Asr	90 590		Gly
	Ile	e Ası	n Vai		r Ile	e Val	Glu	Met 600		s His	s Glr	ı Pro	Asp 605		, Phe	a Asp
	Ile				/ Ala	i Lys	Ser 615		a Glu	ı His	s Val	Asp 620		e Lei	ı GJŻ	/ Ser
	Ala 625	a G1		u Ası	n Asp	Tyr. 630		e Lev	ı Ly:	s Ile	e Ala 635		Gly	ASI	ı Gly	Asp 640
			l Gl	u Pro	0 Arg	g Glr		ı Sei	c Ası	n Lei 650		c Glr	Tr	y Va:	Se:	Gln
	Me	: Pr	o Ph	e Pro 660	o Met							••				



			1133			
<210> 3	*****	erfolkere to a tolue office of the second state		•		·
<211> 2312			,		•	
<212> genor	nic DNA				•	
		rces pombe				
		<u>.</u>				
<400> 3				•		
	CCAAGAAGAG	CAAAACTCAT	AAGAAAAAGA	AAGAAGTCAA		50
				GCTTTGAGTG		100
				ATCAAGAAAA		150
		GAATTTTATA				200
				GTTTTCGACC		250
CTGCTACGTT	AGATAAATTT	GGGAATATGC	TAGGCTCTTC	AGACTTGTTT		300
				ATGCACCTTT		350
		AGTCTCCTAG				400
		GGATATCGAA				450
				TTAATAATTG		500
				TCTATGCTGA		550
AGGCAATGTT	CCTTGACAAG	CAATGCTGGC	TTGAACATTT	AATGCTTGAT		600
AAAAAAACCG	GCTTGGATCC	GAAGGGAATT	AAGCTGCGAG	CAGCTCAGGG		650
GTTTGAAGCA	GCTGATTTTT	TTATCACGGG	CTATTGGATT	TGGAGTAAAG		700
TAATTGAAAA	CCTTGCTGCA	ATTGGTTATG	AGCCTAATAA	CATGTTAAGT		750
				AACGTGATAA		800
ATATTTTTCA	AAGTTAAAAA	TGTTCATTGA	GTACAGCAAC	ATTGTACATA		850
AGAAAAAGGT	AGTGTTGATT	TCTCACTCCA	TGGGTTCACA	GGTTACGTAC		900
TATTTTTTA	AGTGGGTTGA	AGCTGAGGGC	TACGGAAATG	GTGGACCGAC		950
TTGGGTTAAT	GATCATATTG	AAGCATTTAT	AAATGTGAGT	CTCGATGGTT		1000
GTTTGACTAC	GTTTCTAACT	TTTGAATAGA	TATCGGGATC	TTTGATTGGA	1050	
GCACCCAAAA	CAGTGGCAGC	GCTTTTATCG	GGTGAAATGA	AAGATACAGG		1100
TATTGTAATT	ACATTAAACA	TGTTAATATT	TAATTTTTGC	TAACCGTTTT		1150
AAGCTCAATT	GAATCAGTTT	TCGGTCTATG	GGTAAGCAAT	AAATTGTTGA		1200
GATTTGTTAC	TAATTTACTG	TTTAGTTTGG	AAAAATTTTT	TTCCCGTTCT		1250
GAGGTATATT	CAAAAATACA	AATGTGCTCT	ACTTTTTCTA	ACTTTTAATA		1300
GAGAGCCATG	ATGGTTCGCA	CTATGGGAGG	AGTTAGTTCT	ATGCTTCCTA		1350
AAGGAGGCGA	TGTTGTATGG	GGAAATGCCA	GTTGGGTAAG	AAATATGTGC		1400
				ATCAAACAAA		1450
				GATAAGGACC		1500
ACGATGAATT	TGACATAGAT	GATGCATTAC	AATTTTTAAA	AAATGTTACA		1550
GATGACGATI	TTAAAGTCAT	' GCTAGCGAAA	AATTATTCCC	ACGGTCTTGC		1600
				TCTAAATGGA		
				TACTAACCCA		1700
AATAGACTAC	TCTTCCTTAT	GCTCCTGATA	TGAAAATTTA	TTGCGTTCAC		1750
GGGGTCGGAA	AACCAACTGA	GAGAGGTTAT	TATTATACTA	ATAATCCTGA		1800
				AAAGTTGAAA		1850
				TTATTAGGGT		1900
				CCCTTGGTTT		1950
				GCTAATACAA		2000
				TGATCTGAGA		2050
GGAGGACCTC	GCTCGGCAGA	ACACGTCGAT	ATACTTGGAC	ATTCAGAGCT		2100
				CTCTTGAAAT		2150
				CGGTACCAAA		
				AGATTGCAAT		2250
TAACTAACT	ACCGAACAGO	GAAATAATAA	ATGAGATAAA	TCTCGATAAA		2300

CCTAGAAATT AA

्र के देखेल हम के अर्थ केम्पर है जान



<210>~~4

<211> 3685

<212> genomic DNA

<213> Arabidopsis thaliana

<400> 4	mma1 maaa3	************	CACAAAGGAG			50
		AAAGCCGACG				100
		ATGAGGATTC				
		AAATCGAACG				150
	GTTGTTGGTT	CATTGGGTGT	GTGTGTGTAA			200
	CTTTACAACG	CAATGCCTGC		CAGTATGTAA		250
	CACGGGTCCT	TTGCCTGACC	CGCCCGGTGT	TAAGCTCAAA		300
AAAGAAGGTC	TTAAGGCGAA			CTGGGATTGT		350
	CTCGAGCTTT	GGGAAGGCAA		GATGGTTTAT		400
	TTTGTGGGGT	GGAACTTTTG	GTGAAGTCTA	CAAAAGGTGA		450
	TCTCACTCTT	CCTTTATATT	GGGATTTGGA	TTGGATCTGA		500
	CACTTGTTGC	TTCTTCAACA	TCACTCAAAC	TTTAATTCCA		550
	TCTTACTCTT	TACTTTTTTT		GTGAAACGCT		600
ATTTTCTTAA	GAGACTATTT	CTGTATGTGT	AAGGTAAGCG	TTCCAAGGAC		650
GTAATTGGCT	TGGACTATTT	CTGTTTGATT	GTTAACTTTA		•	700
TAGCTGCCTT		GTCATCTTAT	TGCCAAATCT	GTTGCTAGAC		750
	GTCCGTTCAT	AACAAGTTAC		GTCGTTGCGT		800
GTAGATTTAG	CTTTGTGTAG	CGTATAATGA		TATGTTTTGT		850
		CTACATCTGT	GGAAAGTGTG	TTCAGGCTGT		900
GATAGAGGAC	TGTTGCTTTA	TTATTCAACT	ATGTATATGT	GTAATTAAAG		950
CTAGTTCCTT	TTTGATCTTT	CAGCTCAATG	TGCTTTTCTC	AATTTTTTTC	1000	
TCAATTTCAA	AGTTTCACAT	CGAGTTTATT	CACATGTCTT	=		1050
CATCCTCGTT	CTGTTATCCA	GCTTTGAACT	CCTCCCGACC	CTGCTATGGA		1100
TATATTAAAA	AAAAAGTGTT	TTGTGGGTTG	CATCTTTGTT			1150
TCTTCTTCTT	TCGGCTCAGT	GTTCATGTTT	TTGCTATGGT			1200
		CAGTGGTATA		•		1250
		GGCCTCTATG			. 1300	
		GATCCAGCTG			1350	
		CTACTTTGCT			1400	
		CACATATTGG			1450	
ACATGGCTGC		CGGCTTTCGT			1500	
TTCTCATCGT		ATTCTGTTCC			1550	
	CTTAAATATG		and the second s			1600
		GTAATATAGA			1650	
		GTTCCGCATT			1700	
CTACATTTTA			GCTCCTCTGG		1750	
TGGGCCAGAT		AGTATATTAA			1800	
		AAAGCTGTTG			1850	
GCAAAGGATG			ATATCTGCTT		1900	•
		GAACTCAAAG			1950	
		ATCGCTGCAA			2000	
TTGCTGCTTA	. TGTAACTGAA	ACTCTCTTGA	GATTAGACAA	ATGATGAATT	2050	
				GCTTCGACGA		24.52
				GCTATGGAAA		2150
				TTATTCTGCT		2200
				TCTTCTTAAT		2250
TAAAGACTCG	TTGGATTAGT	TGCTCTATTA	GTCACTTGGT	TCCTTAATAT		2300
				AGGATTCTTA		2350
				TGAGAATGAC		2400
				GGTGACACGA		2450
				CTGTTGTGGG		2500
AAAAAGCAAA	AGAACAACGA	AACTTGTGGT	GAAGCAGGTG	AAAACGGAGT		2550
				TTTGGGAAAG		2600
				TTTTCGAGTA		2650
				TGTATGATGA		2700
				TCAGAGTATC		2750
				ACATGGGAAT		2800
TGCTGGGATC	AAAGCTATCG	CTGAGTATAA	GGTCTACACT	GCTGGTGAAG	,	2850

CTATAGATCT	ACTACATTAT	GTTGCTCCTA	AGATGATGCC	GCGTGGTGCC		2300
GCTCATTTCT	CTTATGGAAT	TGCTGATGAT	TTGGATGACA	CCAAGTATCA		2950
AGATCCCAAA	TACTGGTCAA	ATCCGTTAGA	GACAAAGTAA	GTGATTTCTT		3000
GATTCCAACT	GTATCCTTCG	TCCTGATGCA	TTATCAGTCT	TTTTGTTTTC		3050
GGTCTTGTTG	GATATGGTTT	TCAGCTCAAA	GCTTACAAAG	CTGTTTCTGA		3100
GCCTTTCTCA	AAAAGGCTTG	CTCAGTAATA	TTGAGGTGCT	AAAGTTGATA		3150
CATGTGACTC	TTGCTTATAA	ATCCTCCGTT	TGGTTTGTTC	TGCTTTTTCA		3200
GATTACCGAA	TGCTCCTGAG	ATGGAAATCT	ACTCATTATA	CGGAGTGGGG		3250
ATACCAACGG	AACGAGCATA		CTTAACCAGT	CTCCCGACAG	3300	
TTGCATCCCC	TTTCAGATAT	TCACTTCTGC	TCACGAGGAG			3350
	AGCAGGAGTT		ATGGGGATGA			3400
GCTGTCTGAA	CCGGGTACAT					3450
GTCCTAAGTG						3500
ATTCAACCCT						3550
CGCCGGCTAA						3600
GATATCATGG						3650
CGGAGGTAAC				3685		
TTGAATGGTC	GGAGCGTATT	GACCIGAAGC	IGIGA	5005		



<210> 5 <211> 2427 <212> cDNA <213> Arabidopsis thaliana

<400> 5						E0 :
AGAAACAGCT	CTTTGTCTCT (CTCGACTGAT	CTAACAATCC (TAATCTGTG		50
TTCTAAATTC	CTGGACGAGA S	TTTGACAAAG	TCCGTATAGC '	TTAACCTGGT		100
TTAATTTCAA	GTGACAGATA '	IGCCCCTTAT	TCATCGGAAA	AAGCCGACGG		150
AGAAACCATC	GACGCCGCCA '	TCTGAAGAGG	TGGTGCACGA '	TGAGGATTCG		200
CAAAAGAAAC	CACACGAATC '	TTCCAAATCC	CACCATAAGA	AATCGAACGG		250
AGGAGGGAAG	TGGTCGTGCA '	TCGATTCTTG	TTGTTGGTTC	ATTGGGTGTG		300
TGTGTGTAAC	CTGGTGGTTT ·	CTTCTCTTCC	TTTACAACGC	AATGCCTGCG		350
AGCTTCCCTC	AGTATGTAAC	GGAGCGAATC	ACGGGTCCTT	TGCCTGACCC	450	400
GCCCGGTGTT	AAGCTCAAAA	AAAGAAGGTC	TTAAGGCGAA	ACATCCTGTT	450	500
GTCTTCATTC	CTGGGATTGT	CACCGGTGGG	CTCGAGCTTT	GGGAAGGCAA		500 550
ACAATGCGCT	GATGGTTTAT	TTAGAAAACG	TTTGTGGGGT	GGAACTTTTG		
GTGAAGTCTA	CAAAAGGCCT	CTATGTTGGG	TGGAACACAT	GTCACTTGAC		600 650
AATGAAACTG	GGTTGGATCC	AGCTGGTATT	AGAGTTCGAG	CTGTATCAGG		700
ACTCGTGGCT	GCTGACTACT	TTGCTCCTGG	CTACTTTGTC	TGGGCAGTGC		750
TGATTGCTAA	CCTTGCACAT	ATTGGATATG	AAGAGAAAA	TATGTACATG		800
GCTGCATATG	ACTGGCGGCT	TTCGTTTCAG	AACACAGAGG	TACGTGATCA		850
GACTCTTAGC	CGTATGAAAA	GTAATATAGA	GTTGATGGTT	TCTACCAACG		900
GTGGAAAAA	AGCAGTTATA	GTTCCGCATT	CCATGGGGGT	CTTGTATTTT		950
CTACATTTTA	TGAAGTGGGT	TGAGGCACCA	GCTCCTCTGG	GIGGCGGGG	1000	950
TGGGCCAGAT	TGGTGTGCAA	AGTATATTAA	GGCGGTGATG	AACATTGGTG	1000	1050
GACCATTTCT	TGGTGTTCCA	AAAGCTGTTG	CAGGGCTTTT	CTCTGCTGAA		1100
GCAAAGGATG	TTGCAGTTGC	CAGAGCGATT	GCCCAGGAT	TCTTAGACAC		1150
CGATATATTT	AGACTTCAGA	CCTTGCAGCA	TGTAATGAGA	ATGACACGCA		1200
CATGGGACTC	AACAATGTCT	ATGTTACCGA	AGGGAGGTGA	CACGATATGG		1250
GGCGGGCTTG	ATTGGTCACC	GGAGAAAGGC	CACACCTGTT	GTGGGAAAAA		1300
GCAAAAGAAC	AACGAAACTT	GTGGTGAAGC	AGGTGAAAAC	GGAGTTTCCA		1350
AGAAAAGTCC	TGTTAACTAT	GGAAGGATGA	TATCTTTTGG	GAAAGAAGTA		1400
GCAGAGGCTG	CGCCATCTGA	GATTAATAAT	T ATTGATTITC	GAGGTGCTGT		1450
CAAAGGTCAG	AGTATCCCAA	ATCACACCTO	TCGTGACGTG	TGGACAGAGT		1500
ACCATGACAT	GGGAATTGCT	GGGATCAAAC	CTATCGCTGA	GTATAAGGTC		1550
TACACTGCTG	GTGAAGCTAT	AGATCTACTA	A CATTATGTTG	CTCCTAAGAT		1600
GATGGCGCGT	GGTGCCGCTC	ATTTCTCTTI	A TGGAATTGCT	GATGATTIGG		1650
ATGACACCA	GTATCAAGAT	CCCAAATAC	r GGTCAAATCC	GTTAGAGACA		1700
AAATTACCGA	A ATGCTCCTGA	GATGGAAAT	TACTCATTAT	ACGGAGTGGG		1750
GATACCAACC	GAACGAGCAT	ACGTATACA	A GCTTAACCAG	TCTCCCGACA		1800
GTTGCATCC	CTTTCAGATA	TTCACTTCT	G CTCACGAGGA	GGACGAAGAT		1850
AGCTGTCTG	A AAGCAGGAGT	TTACAATGT	G GATGGGGATG	AAACAGTACC		
CGTCCTAAG?	r GCCGGGTACA	TGTGTGCAA	A AGCGTGGCGT	GGCAAGACAA		1900 1950
GATTCAACC	C TTCCGGAATC	AAGACTTAT	A TAAGAGAATA	CAATCACTCT		2000
CCGCCGGCT	A ACCTGTTGGA	. AGGGCGCGG	G ACGCAGAGTG	GTGCCCATGT	•	2050
TGATATCAT	G GGAAACTTTG	CTTTGATCG.	A AGATATCATO	AGGGTTGCCG		2100
CCGGAGGTA	A CGGGTCTGAI	ATAGGACAT	G ACCAGGTCCA	CTCTGGCATA		2150
ጥጥጥር እ አጥርር፣	r · CGGAGCGTAT	TGACCTGAA	G CTGTGAATAT	CATGATCTCT		2200
TTAAGCTGT	C CTGTCAGCTI	ATGTGAATC	C AATACTTTGA	AAGAGAGATC		2250
ATCATCAAT'	T CATCATCATC	GTCATCATC	A TGATGCTCAA	CTCACAAAGA		2300
እር ርርጥር እ ር እ	ል ጥርልጥልርጥጥጥ ር	: GTGCGAAAT	T CTCAATACCI	CTTTAATATT		
ርጥጥ አጥጥር አ አ ነ	Τ GTAAATTATA	CAATCCTAT	C TAATGTTTGA	ACGATAACAC		2350
AAAACTTGC'	T GCNGCCATGT	TTGTTTGTC	T TGTCAAAAGC	ATCAATTTGT		2400
GGGTTAAAA	AAAAAAAA	AAAAAA		2427		

<210> 6		
<211> 671		•
<212> PRT	•	
<213> Arabidopsis thaliana		
	,	
<400> 6		
	QKK PHESSKSHHK KSNGGGKWSC	50
IDSCCWFIGC VCVTWWFLLF LYNAMPA	SFP QYVTERITGP LPDPPGVKLK	100
KEGLKAKHPV VFIPGIVTGG LELWEGK	QCA DGLFRKRLWG GTFGEVYKRP	150
LCWVEHMSLD NETGLDPAGI RVRAVSG	LVA ADYFAPGYFV WAVLIANLAH	200
IGYEEKNMYM AAYDWRLSFQ NTEVRDQ	TLS RMKSNIELMV STNGGKKAVI	250
VPHSMGVLYF LHFMKWVEAP APLGGGG	GPD WCAKYIKAVM NIGGPFLGVP	300
KAVAGLESAE AKDVAVARAI APGELDT	DIF RLOTLOHVMR MTRTWDSTMS	350
MLPKGGDTIW GGLDWSPEKG HTCCGKK	OKN NETCGEAGEN GVSKKSPVNY	400
	KGQ SIPNHTCRDV WTEYHDMGIA	450
GIKAIAEYKV YTAGEAIDLL HYVAPKM		500
PKYWSNPLET KLPNAPEMEI YSLYGVO		550
FTSAHEEDED SCLKAGVYNV DGDETVE		600
	DIM GNFALIEDIM RVAAGGNGSD	650
IGHDQVHSGI FEWSERIDLK L	671	
TRUDANDRI LEMPEKIDDY D	• · -	

<210> 7 <211> 643 <212> cDNA <213> Zea mays <221> CDS <222> (1)..(402)

GAT GAA ACT GTT CCA GTT CTT AGT GCG GGC TAC ATG TGT GCG AAA GGA 96. Asp Glu Thr Val Pro Val Leu Ser Ala Gly Tyr Met Cys Ala Lys Gly 20 25 30

TGG CGT GGC AAA ACT CGT TTC AGC CCT GCC GGC AGC AAG ACT TAC GTG 144
Trp Arg Gly Lys Thr Arg Phe Ser Pro Ala Gly Ser Lys Thr Tyr Val
35 40 45

AGA GAA TAC AGC CAT TCG CCA CCC TCT ACT CTC CTG GAA GGC AGG GGC 192
Arg Glu Tyr Ser His Ser Pro Pro Ser Thr Leu Leu Glu Gly Arg Gly
50 55 60

ACC CAG AGC GGT GCA CAT GTT GAT ATA ATG GGG AAC TTT GCT CTA ATT 240 Thr Gln Ser Gly Ala His Val Asp Ile Met Gly Asn Phe Ala Leu Ile 65 70 75 80

GAG GAC GTC ATC AGA ATA GCT GCT GGG GCA ACC GGT GAG GAA ATT GGT 288 Glu Asp Val Ile Arg Ile Ala Ala Gly Ala Thr Gly Glu Glu Ile Gly 85 90 95

GGC GAT CAG GTT TAT TCA GAT ATA TTC AAG TGG TCA GAG AAA ATC AAA 336 Gly Asp Gln Val Tyr Ser Asp Ile Phe Lys Trp Ser Glu Lys Ile Lys 100 105 110

TTG AAA TTG TAA CCTATGGGAA GTTAAAGAAG TGCCGACCCG TTTATTGCGTTCC 391 Leu Lys Leu 115

AAAGTGTCCT GCCTGAGTGC AACTCTGGAT TTTGCTTAAA TATTGTAATT TTTCACGC 449
TTCATTCGTC CCTTTGTCAA ATTTACATTT GACAGGACGC CAATGCGATA CGATGTTG 507
TACCGCTATT TTCAGCATTG TATATTAAAC TGTACAGGTG TAAGTTGCAT TTGCCAGC 565
TGAAATTGTG TAGTCGTTTT CTTTACGATT TAATANCAAG TGGCGGAGCA GTGCCCCA 623
AGCNAAAAAA AAAAAAAAAA AAAAAAAAAA 643

<210> 8 <211> 115 <212> PRT <213> Zea mays

<400> 8

Arg Glu Lys Ile Ala Ala Leu Lys Gly Gly Val Tyr Leu Ala Asp Gly
1 5 10 15

Asp Glu Thr Val Pro Val Leu Ser Ala Gly Tyr Met Cys Ala Lys Gly
20 25 30

Trp Arg Gly Lys Thr Arg Phe Ser Pro Ala Gly Ser Lys Thr Tyr Val 35 40 45

Arg Glu Tyr Ser His Ser Pro Pro Ser Thr Leu Leu Glu Gly Arg Gly 50 55 60

Thr Gln Ser Gly Ala His Val Asp Ile Met Gly Asn Phe Ala Leu Ile 65 70 75 80

Glu Asp Val Ile Arg Ile Ala Ala Gly Ala Thr Gly Glu Glu Ile Gly 85 90 95

Gly Asp Gln Val Tyr Ser Asp Ile Phe Lys Trp Ser Glu Lys Ile Lys
100 105 110

Leu Lys Leu 115

		,	
			•
sa .			
aagttggagg	ctaacgagaa	tgacnetegg	50
dagecggagg	conttgcacc	carcttcaar	100
gacacgacca	duadcdadcc	tacatctatc	150
acaacccatc	rradatdac	gatcgatacg	200
tanacycatt	atratttra	acusadacas	250
ettteeest:	cctatacaat	aaaaaataaa	300
gcccggggca	22252CCCT	adgyggagga	350
gegggettaa	addidactyt	ggccgagacg	400.
caatccgaga	ggagggccga	acacygegga	450
ggcagaatct	aaacgagtac	acciccaaay	500
acaattgagg	attttattac	tagtaatatt	550
tgaaatttat	gaagagtaat	taaatacggc	
tgactaatta	aaaaaaaatt	ttttttttaa	. 600
	616		
	gacacgacta ggaagccgac acaacgcatc tgaccacggt gtttggggta gcgggctcaa caatccgaga ggcagaatct acaattgagg tgaaatttat	aagttggagg ctaacgagaa gacacgacta ccnttgcacc ggaagccgac ggagcgagcc acaacgcatc tttagatgac tgaccacggt gtgattttgg gtttggggta cctgtgcaat gcgggctcaa aaataaccgt caatccgaga ggagggccga ggcagaatct acaattgagg attttattac tgacaatta gaagagtaat tgactaatta aaaaaaaatt	aagttggagg ctaacgagaa tgacnctcgg gacacgacta ccnttgcacc cagcctcaag ggaagccgac tacatctatc acaacgcatc tttagatgac gatcgatacg tgaccacggt gtgattttgg gcgaaggcga gcttgggata cctgtgcaat aaggggtgga gcgggctcaa aaataaccgt ggtcgagatg caatccgaga ggagggcga atacggcgga ggcagaatct aaacgagtac acaattgag attttattac tagtaatatt tgaaatttat gaagagtaat ttttttctaa 616

<210> 10		•					
<211> 1562							
<212> genon	nic DNA						
<213> Arabi	idopsis tha	aliana					
	-						
<400> 10							
	TATCTTCACA	TTATTCGGTA	GTCATAGCGA	TACTCGTTGT	50		
				GTGTACCCTT		100	
TGATTCTGGT '						150	
				TATATCCGAT		200	
				GCAGCAGTGT		250	
				GTTGTACTAT		300	
				AAACCCGGGT		350	
TCCTCATTTC	GGTTCGACCA	AATCACTTCT	ATACCTCGAC	CCTCGTCTCC		400	
GGTTAGTACT	TTCCAAGATA	TATCATTTTG	GGACATTTGC	ATAATGAACA		450	
AAATAGACAT	AAATTTGGGG	GATTATTGTT	ATATCAATAT	CCATTTATAT		500	
GCTAGTCGGT	AATGTGAGTG	TTATGTTAGT	ATAGTTAATG	TGAGTGTTAT		550	
GTGATTTTCC	ATTTTAAATG	AAGCTAGAAA	GTTGTCGTTT	AATAATGTTG		600	
CTATGTCATG	AGAATTATAA	GGACACTATG	TAAATGTAGC	TTAATAATAA		650	
GGTTTGATTT	GCAGAGATGC	CACATCTTAC	ATGGAACATT	TGGTGAAAGC		700	
				CTAGGAGCTC		750	
CATATGATTT	CAGGTACGGC	CTGGCTGCTT	CGGGCCACCC	GTCCCGTGTA		800	
GCCTCACAGT	TCCTACAAGA	CCTCAAACAA	TTGGTGGAAA	AAACTAGCAG		. 850	
CGAGAACGAA	GGAAAGCCAG	TGATACTCCT	CTCCCATAGC	CTAGGAGGAC		900	
TTTTCGTCCT	CCATTTCCTC	AACCGTACCA	CCCCTTCATG	GCGCCGCAAG		950	
TACATCAAAC	ACTTTGTTGC	ACTCGCTGCG	CCATGGGGTG	GGACGATCTC		1000	
TCAGATGAAG	ACATTTGCTT	CTGGCAACAC	ACTCGGTGTC	CCTTTAGTTA		1050	
ACCCTTTGCT	GGTCAGACGG	CATCAGAGGA	CCTCCGAGAG	TAACCAATGG		1100	
				CGCTTGTCGT		1150	
AACTCCCÇAG	GTTAACTACA	CAGCTTACGA	GATGGATCGG	TTTTTTGCAG		1200	
ACATTGGATT	CTCACAAGGA	GTTGTGCCTT	ACAAGACAAG	AGTGTTGCCT		1250	
				GCATATATGG		1300	
				GGAGGATTCG		1350	
				GGTTAATTTG		1400	
				TAGAGATTGA		1450	
				CTTAAAGAGA		1500	
TTATGAAGCA	GATTTCAATT	ATTAATTATG		TGTTAATGCC		1550	
GTCAATGAAT	GA ·		15	62			

<211> 3896

<210> 11 -

<212> genomic DNA <213> Arabidopsis thaliana <400> 11 ATGGGAGCGA ATTCGAAATC AGTAACGGCT TCCTTCACCG TCATCGCCGT TTTTTTCTTG ATTTGCGGTG GCCGAACTGC GGTGGAGGAT GAGACCGAGT TTCACGGCGA CTACTCGAAG CTATCGGGTA TAATCATTCC GGGATTTGCG 150 TCGACGCAGC TACGAGCGTG GTCGATCCTT GACTGTCCAT ACACTCCGTT 200 GGACTTCAAT CCGCTCGACC TCGTATGGCT AGACACCACT AAGGTCCGTG 250 ATCTTCATTT CCTTCGCTCC TTATTCTGTC GGTCGAGTCA CTTGTTGATG 300 350 CAAATTGGAA GAGCGTGACC TTTACTTTCA CAAGCTCAAG TTAGTCCTTA 1000 1050 1100 TCAGGCTAAT GTCTTTTATC TTCTCTTTTT ATGTAAGATA AGCTAAGAGC TCTGGTCGTC TTCCTTTTTG CAGGTTGACC TTTGAAACTG CTTTAAAACT TCTGGTCGTC TTCCTTTTG CAGGTTGACC TTTGAAACTG CTTTAAAACT
CCGTGGCGCC CCTTCTATAG TATTTGCCCA TTCAATGGTA AATAATGTCT
TCAGATACTT TCTGGAATGG CTGAGGCTAG AAATGCACC AAAACATTAT
TCAGATACTT TCTGGAATGG CTGAGGCTAG AAATGCACC AAAACATTAT
TTCAAGTGCC TTGATCAGCA TATCCATGCT TATTTCCGTG TTGGTACCGG
CCTACTATCC TTAAGTTACC ATTTTATTT TCTCTAATT GGGGGAGTTA 1300
TGTTGTGACT TACTGGATTG AGCTCGATAC CTGATTTTTT GTTGATTTAG 1350
GAGCTCCTCT TCTTGGTTCT GTTGAGGCAA TCAAATCTAC TCTCTCTGGT
GAAACGTTTG CCCTTCTGT TTCTGAGGTG ACCTCTGAT TCTCTCTGGT 1400
GTAAACGTTTG GCCTTCCTGT TTCTGAGGTG ACCTCTGACT TCTCTCTTGT 1450
TTAAAGTAGT TGATATCAAC CAGGTCTTAT AACTCACTGG ATTTTCCTTT 1500
TGAAAGTATT ACTTTTGTTA ATTGAACTGC TGTACGCGAT ATGTATACCT 1550
TAGAACTTTG AGTGCTAATA CAACCAAACC ACATGTACAC TGATTTAGTT 1650
TTCAGAGTATT TATGGTAGAC TTTAAGTTGA ACACTAACCA TGATTTAGTT 1650
TTCAGATTAT TATGGTAGAC TTTAAGTTGA ACACTAACCA TGATTTAGTT 1650
CTTTTTATTT TAATAGGCTA TGATTTTTT ATTGAAACAC ACATGTACAC TGATTTAGTT 1650
GGTTGTTGC CAATTCTTTT GCGTCGTCAT TGATTTGTT ATTGAAAACA TATTCTGTG GGAACTACC TGGACTATT TTTATTT TAATAGGCTA TGATTTCTTT ATTGAAAACA TATTCTGTG GGAACTACC TGGACTATTA TTTGATTGC AGCCTTCAG GGAACTACCT 1800
GGTTGTTGC CAATTCTTTT GCGTCGTCAT TGATGAAACA GAACATTTT CTGGCGGTGC 1900
TGCAAAGAAA GAAACATATTA TAACACACA ACATTCTCC TGGGGGTGC 1900
ACTAGCGGTT AGACTCTAT TATGCACCA TATCCACCAG TAACCACAC CAATGAACA AAAATTTA TAACACTAAC AAAAGTTTCC TTTCATGTC TTTCATGTC TTTCATGAC CAATACCAC AAAAGTTCAC CAATACCACA AAAAGTTCAC CAATACCACA ACAAGTTCAC CAATACCACA AAAAGTTCAC CAATACCACAC AGAAACACA ACAATTCAAC AAAAGTTCAA AAATTCATA AATTCCTTCC TTTCACAGAAC ACAATTCAAC ACAATTCAAC AAAAGTTCAA AAATTCAAC AAAAGTTCAA AAAAGTTCAA AAATTCAAC AAAAGTTCAA AAAAGTTCAA AAATTCAAC AAAAGTTCAA AACAAGTTAAA AACCAATTAA AACCAATTAAA CCGTGGCGC CCTTCTATAG TATTTGCCCA TTCAATGGGT AATAATGTCT 1150 CCTTATTATT GATTATCAGT TCTCTCCTTA TATTATGGAA TGTCTTTTTC GTTTACAGTT ATGAATGCAA AAGGGGGTAT TTTAGTTGAT TGATTCTCTC
ATTCTCTAGT TTGTTTTGAC TAATAGCGTC AATTTTGTTT TTCTAGCAAA
TCTTTGTGAA TTATATATAA CATGCTAACT ATACTTTTCA GGTTGTATCA
TGATGACCCT GTTTTTAATC CTCTGACTCC TTGGGAGAGA CCACCTATAA
AAAATGTATT TTGCATATAT GGTGCTCATC TAAAGACAGA GGTATGATGC
ATTCTCAATA TCACATTATG CGTTGACTTT GTTATTATAT TCCCCATTTG 2650 GTTTGCAATA TCTTTTTGAA TTATGATTTA TCTTCTCCCT TGCATCTTAT GCTATTAAGC GTTAAAGGTA CTAAATGTAT GAAGCTGTCT GTCATAGGTT

	GGTTÄTTÄCT	TTGCCCCAAG	TGGCAAACCT	TATCCTGATA	ATTGGATCAT		Z 900	••
			CTGAAGGTTC		AGGTAATTTT		2950	
	CACGGATATC	ATTTATGAAA		AAGTCTTCTG	TATCAGTCTA		3000 -	
1	CCGCAATGGC	AGAAGTAAAA	CAGGAAGGCA		ACTAAAATTT		3050	
1	GTGGCATGTT	ATCTCAGTTG	CATAAGCAAA	* ****			3100	
	AAGTACTTTT	TTATCATTCC	TTTTGAGCTT		CAGTGGCTTA		3150	
	AAGTGGGAAG	AGGTGTTGCA	TGAAACATGA	CACTTGTATC	AAAGATAACT			
	AGCAAAACAA	AACTAACCCA	TTTCTGAATT	TCATATTATT	AGGAGTAGTC		3200	
	GTGCTTTTAA	AAAATTTGTT	TTAAGAAACC	GAAAAACTAG	TTCATATCTT		3250	
	GATTGTGCAA	TATCTGCAGG	TCTGGAACTG	TGGTTGATGG	GAACGCTGGA	3300		
	CCTATAACTG	GGGATGAGAC	GGTAAGCTCA	GAAGTTGGTT	TTGAAATTAT		3350	
		ACTACTGAAG	ACTAAGATAA	TACTTGCTTC	TGGAACACTG		3400	
	CTTCTTGCAA		ACTGCAATAT		CTACTTTTAT		3450	
	CTTGCTATGT	TCTCTAGTAC					3500	
	TGATTATGAA	ATTGATCTCT			TCCCCAGGTA		3550	
	GCAAGAATTG	GCTCGGACCT				•	3600	
	CTCTTTTTTA	GTTCCTCACC	TTATATAGAT				• • • •	
	CTGGTTATGT	GTTGATTTAC	CTCCAATTTG	TTCTTTCTAA			3650	
	TCTCTGTACT	CCTCAAGAAC	TTGTATTAAT	CTAAACGAGA	TTCTCATTGG		3700	
	GAAAATAAAA	· · · · · · · · · · · · · · · · · · ·	AACACGATGG	AAGCGACGTA	CATGTGGAAC		3750	
	TAAATGTTGA			TCATAGCTAA	CATGACAAAA		3800	
	GCACCAAGGG						3850	
						3896		
	GGGGAAGAGA	ACCGCAGTCI	. GGGRGCIIO					

	<210> 12					•
	<211> 709					
	<212> cDNA					
	<213> Lycope	rsicon escul	entum			
	<400> 12					~ ^
	CTGGGGCCAA	AAGTGAACAT	AACAAGGACA	CCACAGTCAG	AGCATGATGT	. 50
	TCAGATGTAC	AAGTGCATCT	AAATATAGAG			100
	CATTCCCAAT	ATGACAAAGT	TACCTACAAT	GAAGTACATA	ACCTATTATG	150
	AGGATTCTGA	AAGTTTTCCA	GGGACAAGAA	CAGCAGTTTG	GGAGCTTGAT	200
	AAAGCAAATC		TGTCAGATCT	CCAGCTTTGA	TGCGGGAGCT	250
		ATGTGGCATG	ATATTCATCC	TGATAAAAAG	TCCAAGTTTG	300
		TGGTGTCTGA	TCCTCACTAT	TTTCTTCTAT	AAATGTTTGA	350
	GTTTGTATTG	ACATTGTAAG	TATTGCAACA	AAAAGCAAAG	CGTGGGCCTC	400
	TGAGGGATGA		TGGGATTACG	GGAAAGCTCG	ATGTGCATGG	450
•	GCTGAACATT	GTGAATACAG		TCAAATTATA	TTTTGCAAAA	500
	TATTCTCTTT	TTGTGTATTT			CAACGATGCA	550
	GATATGTATT	CGGGGATGTT	CACCTGGGAC		ATTGAAGAGT	600
		ACATCCTGTC			GAAACTTTGT	650
		AACAAGTTTG			AGCGAAATGA	700
	TIGGCGGAAC	MACMAGILIG	CACAAACAII			
	TTCAGAGAG		•	709)	

<210> 13

<211> 623

<212> PRT

<213> Schizosaccharomyces pombe

<400> 13

MASSKKSKTHKKKKEVKSPIDLPNSKKPTRALSEQPSASETQSVSNKSRKSKFGKRLNFILGAILGICGA70 FFFAVGDDNAVFDPATLDKFGNMLGSSDLFDDIKGYLSYNVFKDAPFTTDKPSQSPSGNEVQVGLDMYNE140 GYRSDHPVIMVPGVISSGLESWSFNNCSIPYFRKRLWGSWSMLKAMFLDKQCWLEHLMLDKKTGLDPKGI210 KLRAAQGFEAADFFITGYWIWSKVIENLAAIGYEPNNMLSASYDWRLSYANLEERDKYFSKLKMFIEYSN280 IVHKKKVVLISHSMGSQVTYYFFKWVEAEGYGNGGPTWVNDHIEAFINISGSLIGAPKTVAALLSGEMKD350 TGIVITLNILEKFFSRSERAMMVRTMGGVSSMLPKGGDVAPDDLNQTNFSNGAIIRYREDIDKDHDEFDI420 DDALQFLKNVTDDDFKVMLAKNYSHGLAWTEKEVLKNNEMPSKWINPLETSLPYAPDMKIYCVHGVGKPT490 ERGYYYTNNPEGQPVIDSSVNDGTKVENGIVMDDGDGTLPILALGLVCNKVWQTKRFNPANTSITNYEIK560 HEPAAFDLRGGPRSAEHVDILGHSELNEIILKVSSGHGDSVPNRYISDIQEIINEINLDKPRN 623 <210> 14

<211> 432 .

<212> PRT

<213> Arabidopsis thaliana

<400> 14
MKKISSHYSVVIAILVVVTMTSMCQAVGSNVYPLILVPGNGGNQLEVRLDREYKPSSVWCSSWLYPIHKK70
SGGWFRLWFDAAVLLSPFTRCFSDRMMLYYDPDLDDYQNAPGVQTRVPHFGSTKSLLYLDPRLRDATSYM140
EHLVKALEKKCGYVNDQTILGAPYDFRYGLAASGHPSRVASQFLQDLKQLVEKTSSENEGKPVILLSHSL210
GGLFVLHFLNRTTPSWRRKYIKHFVALAAPWGGTISQMKTFASGNTLGVPLVNPLLVRRHQRTSESNQWL280
LPSTKVFHDRTKPLVVTPQVNYTAYEMDRFFADIGFSQGVVPYKTRVLPLTEELMTPGVPVTCIYGRGVD350
TPEVLMYGKGGFDKQPEIKYGDGDGTVNLASLAALKVDSLNTVEIDGVSHTSILKDEIALKEIMKQISII420
NYELANVNAVNE

<210> 15

<211> 552

<212> PRT

<213> Arabidopsis thaliana

<400> 15

MGANSKSVTASFTVIAVFFLICGGRTAVEDETEFHGDYSKLSGIIIPGFASTQLRAWSILDCPYTPLDFN70
PLDLVWLDTTKLLSAVNCWFKCMVLDPYNQTDHPECKSRPDSGLSAITELDPGYITGPLSTVWKEWLKWC140
VEFGIEANAIVAVPYDWRLSPTKLEERDLYFHKLKLTFETALKLRGGPSIVFAHSMGNNVFRYFLEWLRL210
EIAPKHYLKWLDQHIHAYFAVGAPLLGSVEAIKSTLSGVTFGLPVSEGTARLLSNSFASSLWLMPFSKNC280
KGDNTFWTHFSGGAAKKDKRVYHCDEEEYQSKYSGWPTNIINIEIPSTSARELADGTLFKAIEDYDPDSK350
RMLHQLKKYVPFFVIRNIAHRSSLAGFLLYHDDPVFNPLTPWERPPIKNVFCIYGAHLKTEVGYYFAPSG420
KPYPDNWIITDIIYETEGSLVSRSGTVVDGNAGPITGDETVPYHSLSWCKNWLGPKVNITMAPQILIGKI490
KQQPEHDGSDVHVELNVDHEHGSDIIANMTKAPRVKYITFYEDSESIPGKRTAVWELDKSGY
552

10

Collaboration Collaborati	15	<170	> Pa	tent	In Ve	er. 2	2.0						•		•		
Asp Glu Asn Asn Lys Gly Gly Ser Val His Asn Lys Arg Glu Ser Arg 30 Asn His Tle His His Gln Gln Gly Leu Gly His Lys Arg Arg Arg Gly 35 Tle Ser Gly Ser Ala Lys Arg Asn Glu Arg Gly Lys Asp Phe Asp Arg 50 Lys Arg Asp Gly Asn Gly Arg Lys Arg Trp Arg Asp Ser Arg Arg Leu 65 70 70 80 11e Phe Tle Leu Gly Ala Phe Leu Gly Val Leu Leu Pro Phe Ser Phe 85 85 86 89 89 89 89 89 89 89 89 89 89 89 89 89	20	<211:	> 66 > PR	T T	.romy	ces (cere	<i>r</i> isi	ae								
30 Asn His Ile His His Gln Gln Gly Leu Gly His Lys Arg Arg Arg Gly Ile Ser Gly Ser Ala Lys Arg Asn Glu Arg Gly Lys Asp Phe Asp Arg 50 So So So Rug Lys Arg Trp Arg Asp Ser Arg Arg Leu 65 Arg Asp Gly Asn Gly Arg Lys Arg Trp Arg Asp Ser Arg Arg Leu 80 Ile Phe Ile Leu Gly Ala Phe Leu Gly Val Leu Leu Pro Phe Ser Phe 85 Gly Ala Tyr His Val His Asn Ser Asp Ser Asp Leu Phe Asp Asn Phe 110 Leu Pro Gln Gly Ile Ser Ser Phe Ile Asp Asp Ile Gln Ala Gly Asn 130 Tyr Ser Thr Ser Ser Leu Asp Asp Leu Ser Glu Asn Phe Ala Val Gly 145 Lys Gln Leu Leu Arg Asp Tyr Asn Ile Glu Ala Lys His Pro Val Val 150 Met Val Pro Gly Val Ile Ser Thr Gly Ile Glu Ser Trp Gly Val Ile 180 60 Gly Asp Asp Glu Cys Asp Ser Ser Ala His Phe Arg Lys Arg Leu Trp 205	25		> 1 Gly	Thr	Leu	Phe 5	Arg :	Arg	Asn '	Val	Gln 10	Asn	Gln	Lys	Ser	Asp 15	Ser
Ile Ser Gly Ser Ala Lys Arg Asn Glu Arg Gly Lys Asp Phe Asp Arg So Lys Arg Asp Gly Asn Gly Arg Lys Arg Trp Arg Asp Ser Arg Arg Leu 80 Lys Arg Asp Gly Asn Gly Arg Lys Arg Trp Arg Asp Ser Arg Arg Leu 80 Ile Phe Ile Leu Gly Ala Phe Leu Gly Val Leu Leu Pro Phe Ser Phe 95 Gly Ala Tyr His Val His Asn Ser Asp Ser Asp Leu Phe Asp Asn Phe 100 Val Asn Phe Asp Ser Leu Lys Val Tyr Leu Asp Asp Trp Lys Asp Val 125 Leu Pro Gln Gly Ile Ser Ser Phe Ile Asp Asp Ile Gln Ala Gly Asn 130 Tyr Ser Thr Ser Ser Leu Asp Asp Leu Ser Glu Asn Phe Ala Val Gly 145 Lys Gln Leu Leu Arg Asp Tyr Asn Ile Glu Ala Lys His Pro Val Val 175 Met Val Pro Gly Val Ile Ser Thr Gly Ile Glu Ser Trp Gly Val Ile 180 Gly Asp Asp Glu Cys Asp Ser Ser Ser Ala His Phe Arg Lys Arg Leu Trp		Asp	Glu	Asn	Asti 20	Lys	Gly	Gjà	Ser	Val 25	His	Asr.	Lys	Arg	G1u 30	Ser	Arg
Lys Arg Asp Gly Asn Gly Arg Lys Arg Trp Arg Asp Ser Arg Arg Leu 80 1le Phe Ile Leu Gly Ala Phe Leu Gly Val Leu Leu Pro Phe Ser Phe 95 Gly Ala Tyr His Val His Asn Ser Asp Ser Asp Leu Phe Asp Asn Phe 110 45 Val Asn Phe Asp Ser Leu Lys Val Tyr Leu Asp Asp Trp Lys Asp Val 115 Leu Pro Gln Gly Ile Ser Ser Phe 11e Asp Asp Ile Gln Ala Gly Asn 130 Tyr Ser Thr Ser Ser Leu Asp Asp Leu Ser Glu Asn Phe 145 Lys Gln Leu Lau Arg Asp Tyr Asn Ile Glu Ala Lys His Pro Val Val 175 Met Val Pro Gly Val Ile Ser Thr Gly Ile Glu Ser Trp Gly Val Ile 180 60 Gly Asp Asp Glu Cys Asp Ser Ser Ala His Phe Arg Lys Arg Leu Trp 60 100 Gly Asp Asp Glu Cys Asp Ser Ser Ala His Phe Arg Lys Arg Leu Trp 60 101 Ala Gly Arg Leu Trp 60 102 Arg Leu Trp 60 103 Arg Leu Trp 60 105 Arg Leu Trp 60 106 Gly Asp Asp Glu Cys Asp Ser Ser Ala His Phe Arg Lys Arg Leu Trp 60	30	Asn	His	Ile 35	His	His	Gln	Gln	Gl <u>v</u> 40	Leu	Gly	His	Lys	Arg 45	Arg	Arg	GŢĀ
Lys Arg Asp Gly Asn Gly Arg Lys Arg Trp Arg Asp Ser Arg Arg Leu 80 The Phe Ile Leu Gly Ala Phe Leu Gly Val Leu Leu Pro Phe Ser Phe 95 Gly Ala Tyr His Val His Asn Ser Asp Ser Asp Leu Phe Asp Asn Phe 100 Val Asn Phe Asp Ser Leu Lys Val Tyr Leu Asp Asp Trp Lys Asp Val 115 Leu Pro Gln Gly Ile Ser Ser Phe Ile Asp Asp Ile Gln Ala Gly Asn 130 Tyr Ser Thr Ser Ser Leu Asp Asp Leu Ser Glu Asn Phe 140 Lys Gln Leu Leu Arg Asp Tyr Asn Ile Glu Ala Lys His Pro Val Val 175 Met Val Pro Gly Val Ile Ser Thr Gly Ile Glu Ser Trp Gly Val Ile 180 Gly Asp Asp Glu Cys Asp Ser Ser Ala His Phe Arg Lys Arg Leu Trp 60 Gly Asp Asp Glu Cys Asp Ser Ser Ala His Phe Arg Lys Arg Leu Trp 60 Tyr Ser The Glu Cys Asp Ser Ser Ala His Phe Arg Lys Arg Leu Trp 60 Gly Asp Asp Glu Cys Asp Ser Ser Ala His Phe Arg Lys Arg Leu Trp 60	,	Ile	Ser 50	Gly	Ser	Ala	Lys	Arg 55	Asn	Glu	Arg	Gly	Lys 60	Asp	Phe	Asp	Arg
Gly Ala Tyr His Val His Asn Ser Asp Ser Asp Leu Phe Asp Asn Phe 100 Val Asn Phe Asp Ser Leu Lys Val Tyr Leu Asp Asp Trp Lys Asp Val 115 Leu Pro Gln Gly Ile Ser Ser Phe Ile Asp Asp Ile Gln Ala Gly Asn 130 Tyr Ser Thr Ser Ser Leu Asp Asp Leu Ser Glu Asn Phe Ala Val Gly 145 Lys Gln Leu Leu Arg Asp Tyr Asn Ile Glu Ala Lys His Pro Val Val 165 Met Val Pro Gly Val Ile Ser Thr Gly Ile Glu Ser Trp Gly Val Ile 180 Gly Asp Asp Glu Cys Asp Ser Ser Ala His Phe Arg Lys Arg Leu Trp 205	35			Asp	Gly	Asn	Gly 70	¥≖ā	Lys	yrg	Trp	Arg 75	Asp	Ser	Arg	Arg	80
Gly Ala Tyr His Val His Asn Sar Asp Ser Asp Leu Phe Asp Asn Phe 100 45 Val Asn Phe Asp Ser Leu Lys Val Tyr Leu Asp Asp Tmp Lys Asp Val 115 Leu Pro Gln Gly Ile Ser Ser Phe Ile Asp Asp Ile Gln Ala Gly Asn 135 Tyr Ser Thr Ser Ser Leu Asp Asp Leu Ser Glu Asn Phe Ala Val Gly 155 Lys Gln Leu Leu Arg Asp Tyr Asn Ile Glu Ala Lys His Pro Val Val 165 Met Val Pro Gly Val Ile Ser Thr Gly Ile Glu Ser Tmp Gly Val Ile 180 Gly Asp Asp Glu Cys Asp Ser Ser Ala His Phe Arg Lys Arg Leu Tmp	40	Ile	Phe	Ile	Fen	Gly 85	Ala	Phe	Lau	Gly	Val	Leu	Leu	Pro	Phe	Ser 95	Phe
Leu Pro Gln Gly Ile Ser Ser Phe Ile Asp Asp Ile Gln Ala Gly Asn 130 Tyr Ser Thr Ser Ser Leu Asp Asp Leu Ser Glu Asn Phe Ala Val Gly 150 Lys Gln Leu Leu Arg Asp Tyr Asn Ile Glu Ala Lys His Pro Val Val 165 Met Val Pro Gly Val Ile Ser Thr Gly Ile Glu Ser Trp Gly Val Ile 180 Gly Asp Asp Glu Cys Asp Ser Ser Ala His Phe Arg Lys Arg Leu Trp 205		GŢĀ	Ala	. Tyz	: His	Val	His	Asn	Sar	Asp 105	Ser	Asp	Leu	Phe	Asp 110	Asn	Phe
Tyr Sar Thr Ser Ser Leu Asp Asp Leu Ser Glu Asn Phe Ala Val Gly 150 Lys Gln Leu Leu Arg Asp Tyr Asn Ile Glu Ala Lys His Pro Val Val 165 Met Val Pro Gly Val Ile Sar Thr Gly Ile Glu Ser Trp Gly Val Ile 180 Gly Asp Asp Glu Cys Asp Ser Ser Ala His Phe Arg Lys Arg Leu Trp 205	45	Val	Asn	Phe 113	e Asç	ser	Leu	Lys	Val 120	Tyr	Leu	. Asp	Asp	T=p 125	Lys	Asp	Val
Tyr Ser Thr Ser Ser Leu Asp Asp Leu Ser Glu Asn Phe Ala Val Gly 145 Lys Gln Leu Leu Arg Asp Tyr Asn Ile Glu Ala Lys His Pro Val Val 165 Met Val Pro Gly Val Ile Ser Thr Gly Ile Glu Ser Trp Gly Val Ile 185 Gly Asp Asp Glu Cys Asp Ser Ser Ala His Phe Arg Lys Arg Leu Trp 205		Leu	Pro 130	Gl:	gly	, Ila	Ser	Ser 135	Phe	lle	Asp	As;	Ile 140	Gla	Ala	Gly	ASD
Met Val Pro Gly Val The Ser Thr Gly The Glu Ser Trp Gly Val The 180 185 190 60 Gly Asp Asp Glu Cys Asp Ser Ser Ala His Phe Arg Lys Arg Leu Trp 205	50	Ty: 143	: Sa:		r Ser	: Ser	. Len	. Asī	Asp) Lau	. Ser	Glu 155	i Asn	Phe	. Als	. Val	. Gly 150
180 133 60 Gly Asp Asp Glu Cys Asp Ser Ser Ala His Phe Arg Lys Arg Leu Trp	55	ŗķs	s Gli	a Le	: Le	1 Arc	. Asp	Ty:	: Ast	: Ile	Glu 170	: Als	r Tha	: His	? Pro	val 17	Val
60 Gly Asp Asp Glu Cys Asp Ser Ser Ala His Phe Arg Lys Arg Leu Trp 195 200		Me	. Va.	1 ?~	c Gly	y Val	Ile	s Sec	ב מלב	Gly 185	n Ile	e Gļ	: Se¤	. <u> </u>	190 190	· Val	L Ils
	60	Gly	y As	p As 19	p Gl ^e S	n CĀi	s Asī	Se:	r Sec 200	r Ala	L Hi	s Ph	e Arq	20:	a Arq	; Le	ı TIP

	Gly Ser Phe Tyr Met Leu Arg Thr Met Val Met Asp Lys Val Cys Trp 210 215
5	Leu Lys His Val Met Leu Asp Pro Glu Thr Gly Leu Asp Pro Pro Ash 240
	Phe Thr Leu Arg Ala Ala Gln Gly Phe Glu Ser Thr Asp Tyr Phe 11e 255
10	Ala Gly Tyr Trp Ile Trp Asn Lys Val Phe Gln Asn Leu Gly Val Ile 270 265
15	Gly Tyr Glu Pro Asn Lys Met Thr Ser Ala Ala Tyr Asp Trp Arg Leu 285 275 280 285 Lys Leu Lys
)	Ala Tyr Leu Asp Leu Glu Arg Arg Asp Arg Tyr Phe Thr Lys Leu Lys 300 290 295 300 290 295
20	Glu Gln Ile Glu Leu Phe His Gln Leu Ser Gly Glu Lys Val Cys Leu 320 315 305
25	Ile Gly His Ser Met Gly Ser Gln Ile Ile Phe Tyr Phe Met Lys Trp 335 325 327 Cly Gly Arg Gly Trp Val
23	Val Glu Ala Glu Gly Pro Leu Tyr Gly Asn Gly Gly Arg Gly Trp Val 345 340 340 340 340 340 340 340 340 340
. 30	Asn Glu Kis Ile Asp Ser Phe Ile Asn Ala Ala Gly Thr Leu Leu Gly 365 355 360 365 Ala Pro Lys Ala Val Pro Ala Leu Ile Ser Gly Glu Met Lys Asp Thr 375
ŕ	
35	- no
)4	Arg Ile Glu Arg Val Lys Met Leu Gln Thr Trp Gly Gly Ile Pro Ser 415 405 0
)**	Met Leu Pro Lys Gly Gill Gid 425
4	Ser Glu Asp Ala Leu Asn Asn Asn Thr Asp Thr Tyr Gly Asn Phe Ile 445 445
	Arg Phe Glu Arg Asn Thr Ser Asp Ala Phe Asn Lys Asn Leu Thr Met 450 450 450 450 450 450 450 450 450
;	Lys Asp Ala Ile Asm Met Thr Leu Ser Ile Ser Pro Glu Trp Leu Gln 480 465 470 480 480
	Arg Arg Val His Glu Gln Tyr Ser Phe Gly Tyr Ser Lys Asn Glu Glu Arg Arg Val His Glu Gln Tyr Ser Phe Gly Tyr Ser Lys Asn Glu Glu Glu Arg Arg Val His Glu Gln Tyr Ser Phe Gly Tyr Ser Asn Pro Met
	55 Glu Leu Arg Lys Asn Glu Leu His His Lys His Trp Ser Asn Pro Mer 510 500
	Glu Val Pro Leu Pro Glu Ala Pro His Met Lys Ile Tyr Cys Ile Tyr 60 515

	ly Val Asn Asn Pro Thr Glu Arg Ala Tyr Val Tyr Lys Glu Glu Asp 530 535	
5	sp Ser Ser Ala Leu Asn Leu Thr Ile Asp Tyr Glu Ser Lys Gln Pro 560 45	
	al Phe Leu Thr Glu Gly Asp Gly Thr Val Pro Leu Val Ala His Ser 575 565	
10	tet Cys His Lys Trp Ala Gln Gly Ala Ser Pro Tyr Asn Pro Ala Gly 580 585	
1.5	lle Asn Val Thr Ile Val Glu Met Lys His Gln Pro Asp Arg Phe Asp 595 600 71. Law Clin Ser	
15	The Arg Gly Gly Ala Lys Ser Ala Glu His Val Asp The Leu Gly Ser 610 615 620	
20	Ala Glu Leu Asn Asp Tyr Ile Leu Lys Ile Ala Ser Gly Asn Gly Asp 625 630 640	
	Leu Val Glu Pro Arg Gln Leu Ser Asn Leu Ser Gln Trp Val Ser Gln 655 655	
25	Met Pro Phe Pro Met 660	
30	<210> 2Q <211> 387 <212> PRT <213> Arabidopsis thaliana	
35	<pre><400> 2 Val Gly Sar Asn Val Tyr Pro Leu Ile Leu Val Pro Gly Asn Gly Gly 15 1</pre>	
	Asn Gln Leu Glu Val Arg Leu Asp Arg Glu Tyr Lys Pro Ser Ser Val 25 30	
40	Trp Cys Ser Ser Trp Leu Tyr Pro Ile His Lys Lys Ser Gly Gly Trp 45	
4:	Phe Arg Leu Trp Phe Asp Ala Ala Val Leu Leu Sar Pro Phe Thr Arg 50 55	
	Cys Phe Sar Asp Arg Met Met Leu Tyr Tyr Asp Pro Asp Leu Asp Asp 80 65	e*
5	Tyr Gln Asn Ala Pro Gly Val Gln Thr Arg Val Pro His Phe Gly Ser 85	
	The Lys Ser Leu Leu Tyr Leu Asp Pro Arg Leu Arg Asp Ala The Ser 100 100	
5	Tyr Met Glu His Leu Val Lys Ala Leu Glu Lys Lys Cys Gly Tyr Val 125	
	Ash Asp Gin Thr Ile Leu Gly Ala Pro Tyr Asp Phe Arg Tyr Gly Leu 130 140	

	Ala 145	Ala	a S	er	Gly	His	P= 15	o S	Ser	Arg	, V	al .	Ala	Se 15	r (Sln	Phe	Leu	Gli	1 As 1	50 60
5	Leu	Ly	s C	iln	Leu	Vai 165	G1	u I	ŗĀ2	Thi	r S	Ser	Ser 170	G1	ų i	Asn	Glu	Gly	Ly: 17	s P: 5	ro
	Val	Il	e I	Leu	Leu 180	Ser	н:	is	Ser	Let	u (31y 185	Gly	Ĺ€	eu :	Phe	Val	Leu 190	Hi	s P	he
1Ò	Leu	As	m. i	Arg 195	Thr	The	: P:	ro	Ser	T17	р і 0	yzā	Arg	Ľ	/S	Tyr	Ile 205	ŗăs	Hi.	s P	he
1.5		21	LO						213							220	Gln				
15	225						2	30						_	33		Asn				
) 20						24	5						۱ ل ک	J			TIP				
					26	3						202					Val				
25				275	•					ا کد	5 0										Ile.
30		2	90						29:	2						300					
30	30.	5 ·						310						•	ں نے ر		Cy:				
35						3	25										/ Ly				
		•			34	10						34.	_				o Gl	-			
):40				35	5					3	6 O U	ı			•			-			Glu
45	=1	:e	As:	G1	y V	al S	er	Hi	s Th	12 S	se:	: Il	e L	eu	Lys	38 38	0 D G1	u I	le i	li.	Lèu
45		/s 35	Glı	ı Il	e														•		
50	<: <:	211 212	> : > :	3 Q 3 8 9 PRT Amair	nido	psi	= =	hel	ian	2								•			
55	1	eu 1		⊑ Ľ\			5														Pro
60	G	ly	Il	e Va	ll T	<u>h=</u> 20	Gly	G1	y L	e:	G1:	u Le 2	eu T 25	<u></u>	G1	u Gl	y Ly	/s G	30	CĀ2	Ala

	Asp	Gly	ŗ L	eu 35	Phe	A≝g	ĽУ	s A	æg 1	Leu 40	TI	. p @	;ly	Gly	Thr	Phe 45	L	eu (Cys '	Tr	Ç
5	Val	Gli 50) 1	is	Met	Ser	· Le	u A	.sp 55	Asn	G!	Lu 7	rhr	Gly	Бел 60	As	ב כ	ro.	Ala	G1;	Ā
	Ile 65	Ar	g V	/al	Arg	Ala	v Va	ll 5	Ser	GŢĀ	L	eu 1	Val	Ala 75	Ala	Ası	Ţ	ζĀΞ	Phe	Al 8	<u>a</u> 0
10	Pro	G1	y T	īĀī	Phe	Va:	נית ז 5	י קר	Ala	Val	L	eu	Ile 90	Ala	Asn	Le	n i	Ala	His 95	IJ	.e
4.4					100						_	0.5			Tyr						
15				115						120	,				. Ser						
20		13	0						T72						140						
	145	;					. 1	.50													
25						16	55						. 11	,	/ Gly						
70					18	0					•	.00			y Gl						
30	Va.	1 P	20	Lys 199	s Al	a V	<u> 1</u>	Ala	Glā	те 20	u : 0	Phe	Se	<u> </u>	a Gl	u A	1= 05	Lys	Asţ	M	let
35		2	10						215	•					r Me 22	•					
	22	5						230													
40						2	45						2.0		y Va						
45					2:	50						200	,		r Pr			•	_		
45	P=	:o 1	Phe	Gl 27	n I. '5	Le E	he	Thi	: Se	r Ai 28	Lz 30	His	s G2	.u G1	u As	;p (:1u :85	As:	se Se	r (Cys
50	Le	eu l	Ŀys 290	; <u>al</u>	.a G	1y (/al	Ty:	29	n Va 5	21	Ası	G]	ly As	ي و: 3 (lu 1	hr	Va.	l Pr	Э· °	Va <u>l</u>
	L:	eu :	Sec	: Al	.a G	ly :		Me:	c Cy	s A	la	Ly	s Al	la T:	ry Ai L5	rg (lly	· Ly	s Th	=	Arg 320
55						:	325		e Ly				33	30	rg G				نہ د	3	
. -		<u>-0</u>	Pr	s Al	ia A 3	sn. : 40	ren	Le	u Gl	u G	ŀY	A_==	ç G: 5	Ly·T	1 <u>≭</u> G	ln s	Ser	: Gl:	y Al O	.£	His
60	V	al	As;	9 I:	le M	at :	Gly	As	n Ph	le A	<u>l</u> a	Le	1 I	le G	li A	50 :	<u> </u>	Me	ت۾ ت	5	V <u>al</u>

	365	
	Ala Ala Gly Gly Asn Gly Ser Asp Ile Gly His Asp Gln Val His Ser	
	370 375	
5	Gly Ile Phe Glu Trp 385	.'
10	<210> 4 Q <211> 1986 <212> DNA <213> Saccharomyces cerevisiae	
15)	<220> <221> CDS <222> (1)(1983)	
20	<400> 4 atg ggc aca ctg ttt cga aga aat gtc cag aac caa aag agt gat tct 48 atg ggc aca ctg ttt cga aga aat gtc cag aac caa aag agt gat tct 48 Met Gly Thr Leu Phe Arg Arg Asn Val Gln Asn Gln Lys Ser Asp Ser 15 10	: - -
25	gat gaa aac aat aaa ggg ggt tot gtt cat aac aag cga gag agc aga 96 gat gaa aac aat aaa ggg ggt tot gtt cat aac aag cga gag agc aga 96 Asp Glu Asn Asn Lys Gly Gly Ser Val His Asn Lys Arg Glu Ser Arg 20 20	•
3(aac cac att cat cat caa cag gga tta ggc cat aag aga aga agg ggt. 1 Asn His Ile His His Gln Gln Gly Leu Gly His Lys Arg Arg Arg Gly 40 45	44
2(att agt ggc agt gca aan agz aat gag cgt ggc aan gat ttc gac agg Ile Ser Gly Ser Ala Lys Arg Asn Glu Arg Gly Lys Asp Pne Asp Arg 55 60	.92
. 3	aaa aga gac ggg aac ggt aga aaa cgt tgg aga gat tcc aga aga ctg aaa aga gac ggg aac ggt aga aaa cgt tgg aga gat tcc aga aga ctg Arg Arg Asp Gly Asn Gly Arg Lys Arg Tro Arg Asp Ser Arg Arg Leu Lys Arg Asp Gly Asn Gly Arg Lys Arg Tro Arg Asp Ser Arg Arg Leu 70	240
)4	65	288
	ggc gct tat cat gtt cat aat agc gat agc gac ttg ttt gac aac ttt Gly Ala Tyr His Val His Asn Ser Asp Ser Asp Leu Phe Asp Asn Phe 105	336
	gta eat tit gat toe cit aas gtg tat tig gat gat tgg aas gat gtt Val Asn Phe Asp Ser Leu Lys Val Tyr Leu Asp Asp Tro Lys Asp Val 120 125	384
• •	ctc cca caa ggt ata agt tog tit att gat gat att cag gct ggt aac Leu Pro Gln Gly Ile Ser Ser Phe Ile Asp Asp Ile Gln Ala Gly Asn 135	432
	tac too aca tot tot the gat gat one ago gas ast the god get ggt tac too aca tot tot the gat gat one ago gas ast the god get ggt tac too aca tot tot the gat gat one ago gas ast the god get ggt tac too aca tot tot the gat gat one ago gas ast the god get ggt tac too aca tot tot the gat gat one ago gas ast the god get ggt tac too aca tot tot the gat gat one ago gas ast the god get ggt tac too aca tot tot the gat gat one ago gas ast the god get ggt tac too aca tot tot the gat gat one ago gas ast the god get ggt tac too aca tot tot the gat gat one ago gas ast the god get ggt tac too aca tot tot the gat gat one ago gas ast the god get ggt tac too aca tot tot the gat gat one ago gas ast the god get ggt tac too aca tot tot the gat gat one ago gas ast the god get ggt tac too aca tot tot the gat gat one ago gas ast the gat gat gat one ago gat gat one ago gat gat gat one gat gat gat one gat	480
	60 eas cae ctc tra cgt get tat ast atc gag gcc ass cat cct gtt gta Lys Gln Leu Leu Arg Asp Tyr Asn Ile Glu Ala Lys His Pro Val Val	528

28/53

-					165					17	0					175			
5	atg gi Met Va	tt c	520 (ggt Sly 180	gtc Val	att Ile	tct Ser	acg Thr	gga Gly 185		t ga e G	aa a lu S	egc Sez	TIP Tgg	gga Gly 190	gtt Val	at Il	t Le	576
	GIY A	sp 1	jat Asp 195	G]n àsa	tgc Cys	gat Asp	agt Ser	tct Ser 200	gcg	ca Hi	t t s P	tt d	egt Arg	aaa Lys 205	CGG Arg	Leu	t T	tb Ba	624
10	gga a Gly S	gt : er :	ttt Phe	tac Tyr	atg Met	ctg Leu	aga Arg 215	aca Thr	at c Met	gt Va	t a	tg (let :	gat Asp 220	aaa Lys	gtt Val	tgt Cys	T	rb aa	67:2
15	ttg a Leu I 225	-	cat His	gta Val	atg Met	tta Leu 230	gat Asp	CCT	ga: Gl:	a ao 1 Tì	-	igt 235	ctg Leu	gac Asp	cca Pro	Pr	g a o A 2	ac sn :40	720
20	ttt a	acg Thr	ctz Leu	cgt	gca Ala 245	ATS	Gln	ggc Gly	tt: Ph		aa t lu : 50	ca Ser	act Thr	gat Asp	Tyr	tt Ph 25	c a e I 5	itc le	768
25	gca (ejy aaa	tat Tyr	tgg Trp 260	ITE	tgg	aac Asi	aaa Lys	gt Va 26		tc : he :	caa Gln	aat	ctg Leu	ggs Gly 270	gt Va V	a a	itt Ile	815
	GJ⊼ āāc	tat Tyr	gaa Glu 275	Pro	aat Ast	aaa Lys	ato Me	g accept The	- 5	t g	ct la	gcg Ala	tat	gat Asr 285		g aç Ç Az	g (ctt Leu	864
30	gcz Alz	tat Tyr 290	tta Leu		t cta p Lev	a gai	ag 1 Air	يسمي	g As	ic s g J	ra aa	tac Tyr	Phe 300		g aa Ly	g ci s L	eu Eu	aag Lys	912
35	gaa Glu 305			e Gi	a cto u Le	g tt u Ph 31	e Hr	t ca s Gl	a ti n L	eu S	igt Ser	ggt Gly 315		a aa 1 Ly	e gt s Va	t to	ğt YS	tta Leu 320	960
40		gga Gly	r Car	tc S Se	t at r Me 32	g gg t Gl 5	t to y Se	t ca r Gl	g 2 n I	_=	atc Ile 330	ttt Phe	ta Ty	c ti r Ph	t at e Me	ga E L 3	aa Ys 35	tgg Trp	1008
45	gtc Val	gaç Glu	<u>. Al</u>	a Gl	a gg u Gl	À BR	t ct	et ta	(= G	gt 1 <u>y</u> 45	aat Asn	G17	gg Gl	t cg y Ar		je t SO	rb âĉ	gtt Val	1056
:	aac Asn	gaa G1:		c at		it to sp Se	a to	ne L.	et a le A 50	at sn	gca Ala	gca	a gg a Gl	y Th		eu L	tg eu	GJĀ	1104
50		. cc:	c Ly	s Y] G GO	a gt La Va	it ca	CO A	ct c la L 75	ta a eu I	:: :le	agt Ser	G1;	r ga y Gl 39		g æ	ee c	at Sp	acc Thr	1152
55	2.5 11e 385	: ca 4 Gl:		a a nı A	ic ac sn Ti	eg.t: nr L: 3:	ia c au A	cc a la M	tig t et T	at Y=	GŢĀ	to Let		e ad u Ly	ig t /s ?	to the S	tc he	Ser 400	1200
60			t ga e Gl	ig a	rg Va	ca a al 1:	aa a Ys M	tg I et L	ta c eu C	iaa iln	209 Thr 410	;	g gg	r gg .y G	jt a Ly I	ta d le i	220 220 115	cca Ser	1248

	atg cta cc Met Leu Pr	a aag gga o Lys Gly 420	a gaa ga / Glu Gl	g gtc u Val	att tg Ile Tr 425	b ell e aea	gat atg Asp Met	aag to Lys Se 430		1296
5	tca gag ga Ser Glu As 43	D WIS TE	g aat aa u Asn As	c aac n Asn 440	****	c aca p Thr	tac ggc Tyr Gly 445	aat t Asn P	tc att he Ile	1344
10	cga ttt ga Arg Phe Gl 450	la agg aa Lu Arg As	n ant se	ge gat er Asp 55	gct the	cc aac ne Asn	aaa aat Lys Asi 460	ttg a Leu T	ca atg hr Met	1392
15	aaa gac go Lys Asp Ai	la Ile As	470	ii nec	. 502 -	475			480	1440
) 20	aga aga g Arg Arg V	al His G	ag cag t lu Gln T 85	ac tco yr Sei	g ttc g r Phe G 4	gc tat ly Tyr 190	tcc aa Ser Ly	g aat g s Asn (gaa gaa 31u Glu 195	1488
	gag tta a Glu Leu A	ga aaa a rg Lys A 500	at gag o sn Glu I	ta ca Jeu Hi	c cac a s His I 505	ag cad	tgg to Trp Se	g aat (r Asn) 510	cca atg Pro Met	1536
25	gaa gca q Glu Val		ca gaa q ro Glu i	gct cc Ala Pr 52	··	atg aa: Met Ly:	a atc ta s Ile Ty 52	t tgt r Cys 25	eta tec Ile Tyr	1584
30	ggg gtg : Gly Val . 530		Sto lum	gaa ag Glu Ar 535	g gca	tat gt Tyr Va	a tat as 1 Tyr Ly 540	ag gaa ys Glu	gag gat Glu Asp	1632
35	gac tcc Asp Ser 545	Ser Ala	Leu Asn	Hen T		55			560	1680
)40	Val Phe	ctc acc Leu Thr	565		-:	570			575	1728
	atg tgt Met Cys	cac aaa His Lys 580	tgg gcc Trp Ala	cag g Gln G	gt gct ly Ala 585	tca co Ser P	eg tac a ro Tyr A	ac cct Lsn Pro 590	gcc gga Ala Gly	1776
45	att aac Ile Asn	git act Val Thr 595	att gtg Ile Val	Gran	itg aaa Met Lys Moo	cac c His G	ag cca (ln Pro l	yat cga Asp Arg 605	ttt gat Phe Asp	1824
50	ata cgt Ile Arg 610	Gjā Gjā Gār Gās	gca aaa Ala Lys	agc 9 Ser 1	ycc gaa Ala Glu	cac g His V	ta gac al Asp 620	atc ctc Ile Lev	ggc agc	1872
5:			gat tac Asp Tyr 630	775	ttg æææ Leu Lys	att g : Ile A 6	ca ago la Ser	ggt ææt Gly Asi	ggc gat Gly Asg 640	1920
6	ctc gtc Leu Val	: gag cca . Glu Pro	ogo caa Arg Gli 645	tes :	tct aat Ser Asr	teu S 650	er Gju	tgg gt: Try Val	tet caq I Ser Gli 655	1963

atg ccc ttc cca atg taa Met Pro Phe Pro Met 660

<210> 5 a <211> 661 <212> PRT <213> Saccharomyces cerevisiae 10 Met Gly Thr Leu Phe Arg Arg Asn Val Gln Asn Gln Lys Ser Asp Ser Asp Glu Asn Asn Lys Gly Gly Ser Val His Asn Lys Arg Glu Ser Arg 15 Asn His Ile His His Gln Gln Gly Leu Gly His Lys Arg Arg Arg Gly Ile Ser Gly Ser Ala Lys Arg Ash Glu Arg Gly Lys Asp Phe Asp Arg
50 55 20 Lys Arg Asp Gly Asn Gly Arg Lys Arg Trp Arg Asp Ser Arg Arg Leu 65 70 75 25 The Phe The Leu Gly Ala Phe Leu Gly Val Leu Leu Pro Phe Ser Phe Gly Ala Tyr His Val His Asn Ser Asp Ser Asp Leu Phe Asp Asn Phe 30 105 Val Asn Phe Asp Ser Leu Lys Val Tyr Leu Asp Asp Trp Lys Asp Val 35 Leu Pro Gln Gly Ile Ser Ser Phe Ile Asp Asp Ile Gln Ale Gly Asn Tyr Ser Thr Ser Ser Leu Asp Asp Leu Ser Glu Asn Phe Ala Vai Gly 40 Lys Gln Leu Leu Arg Asp Tyr Asn Ile Glu Ale Lys His Pro Val Val Met Val Pro Gly Val Ile Ser Thr Gly Ile Glu Ser Trp Gly Val Ile Gly Asp Asp Glu Cys Asp Ser Ser Ala His Phe Arg Lys Arg Leu Trp 195 50 Gly Ser Phe Tyr Met Leu Arg Thr Met Vel Met Asp Lys Val Cys Trp Den Lys His Val Met Leu Asp Pro Glu Thr Gly Leu Asp Pro Pro Asn 230 55 Phe Thr Leu Arg Ala Ala Gin Gly Phe Glu Ser Thr Asp Tyr Phe Ila Ala Gly Tyr Trp Ile Trp Asn Lys Val Phe Gln Asn Leu Gly Val Ile

SUBSTITUTE SHEET (RULE 26)

60

•					•											
	Gly	Tyr	Glu 275	Pro	Asn	Lys	Met '	Thr 280	Ser	Ala	Ala	Tyr	Asp 285	qrr	Arg	Leu
5	Ala	ту <u>т</u> 290	Leu	qaA	Leu	GIu	Arg 295	Arg	qeA	Arg	Tyr	Phe 300	Thr	Lys	Leu	Lys
10	Glu 305	Gln	Ile	Glu	Lеи	Phe 310	His	G1n	Leu	Ser	Gly 315	Glu	Lys	Val	Cys	Leu 320
10	Ile	Gly	His	Ser	Met 325	Gly	Ser	Gln	Ile	Ile 330	Phe	ŢŸĭ	Phe	Met	Lys 335	Tro
15	Val	Glu	Ala	Glu 340	Gly	Pro	Leu	Tyr	Gly 345	Asn	Gly	Gly	Arg	Gly 350	Trp	Val
)	Asn	Glu	His 355	Ile	Asp	Ser	Phe	Ile 360	Asn	Ala	Ala	Gly	Thr 365	Lėu	Leu	Gly
20	Ala	Pro 370	Lys	Ala	Val	Pro	Ala 375	Leu	Ile	Ser	Gly	G1u 380	Met	Lys	Asp	Thr
25	Ile 385	Gln	Leu	Asn	Thr	Leu 390	Ala	Met	Tyr	G17) Leu	Glu	Lys	Phe	Phe	Ser 400
	Arg	Ile	Glu	Arg	Val 405	Lys	Met	Leu	Gln	Thr 410	Tro	Gly	Gly	Ile	Pro 415	Ser
30	Met	Leu	Pro	Lys 420		Glu	Glu	Val	11e 425	Trp	Gly	Asp	Met	Lys 430	Ser	Ser
	Ser	Glu	Asp 435		Leu	Asn	Asn	Asn 440	Thr	Asp	Thr	Tyr	Gly 445	Asn	Phe	Ile
35	yzg	Phe 450		. Arg	Asn	Thr	Se r 455	Asp	Ala	Phe	Asn	Lys 460	Asn	. Leu	Thr	Met
)40	Lys 465		Ala	. Ile	Asn	Mec 470		Lev	Ser	Ile	8e= 475	Pro	Glu	TTP	Leu	Gln 480
. , ,	Arg	Arg	Val	His	Glu 485		. Tyr	Ser	Phe	Gly 490		Ser	. Lys	Asn	495	Glu
45	Glu	Leu	Arg	500		Glu	Leu	. His	His 505		: His	TIL	Ser	510	Pro) Met
	Glu	Val	. Pro		Pro	Glu	Ala	9ro 520		Met	Lys	Ile	1771 525	Cys	: Ile	Tyr
5 0	GŢĀ	Val		Asn	Pro	The	Glu 535		Ala	Ty	. Val	Ty: 540	Lys	s Glu	Glu	ı Asp
55	Asp 545		Sez	: Als	. Leu	ASS 550		The	: Ila	e Asg	Ty:	Glu	Sex	· Lys	Glr	560
,,	Val	Phe	: Lev	ı The	Glu 565		Asp	Gly	Th	Val 570	. Pro	Leu	val	Ala	His 575	Ser
60	Met	Cys	His	: Lys 580		Ale	Glo	Gly	Ala 585	Sez	Pro	Туг	· Asi	9 PFC) Ala	e Gly

			595	Thr				900								
5	Ile	Arg 610	Gly	Gly	Ala	Lys	Ser 615	Ala	Glu	Hìs	Val	Asp 620	Ile	Leu	GŢĀ	Ser
-	Ala 625	Glu	Leu	Asn	Asp	Tyr 630	Ile	Leu	Lys	Ile	Ala 635	Ser	Gly	Asn	Gly	Asp 640
10	Leu	Val	Glu	Pro	AIG 645	Gln	Leu	Ser	Asn	Leu 650	Ser	Gln	Trp	Val	ser 655	Glr
	Met	Pro	Phe	Pro 660											-	
15																

```
SÉQUENCE LISTING
<110> Stymme Dr., Sten
<120>
<110>
 < 14 U >
 <141>
 <150>
 <210> 1 b
 <211> 1986
 <212> genomic DNA
 <213> Saccharomyces cerevisiae
 <220>
) <221> CDS
 <2225 (1) .. (1983)
  <4000 I
  atg ego aca ong the oga aga aat gho ong sac cas ang agh gan non
  Met Gly Thr Lou Phe Arg Arg Asn Val Gln Asn Gln Lys Ser Asp Sor
                                        10
  gat gan and ant ann ggg ggt tot gtt cat and mag ogn gag age agn
  Asp Glu Asn Asn Lys Gly Gly Ser Val His Asn Lys Arg Glu Ser Arg
                                    25
               20
  acc cac att cat cat cas cag ggs tte ggc cat asg aga aga agg ggt
                                                                      144
  Asn His Ile Ris His Gln Gln Gly Leu Gly His Lys Arg Arg Arg Gly
                                40
            35 -
   att agt ggc agt gca att agt aat gag cgt ggc aat gkt ttc gkc agg
                                                                       192
   The Ser Gly Ser Ala Dys Arg Asn Glu Arg Gly Dys Asp Pho Asp Arg
                            55
        50
   sat aga gad ggg aad ggt aga aaa ogt ogg aga gat too aga aga otg
                                                                       240
   Ly: Arg Asp Gly Asn Gly Arg Dys Arg Trp Arg Asp Ser Arg Arg Deu
                                             75
                         70
    £5
   latt too abt out ggt gos the tha ggt gos che the ore the ago tet
    lie Pho Ile Leu Gly Ala Phe Leu Gly Val Leu Leu Pro Phe Ser Phe
                                          90
                     25
    SET GOT THE CAR GOT CAR AND AGO GAT AGO GRE TOG THE GAO ARE THE
    Gly Ala Tyr Mis Val Mis Asn Ser Asp Ser Asp Leu Phe Asp Asn Phe
                                                         110
                                     103
                100
```

gta aat tit gat toa ott 222 gtg tat tig gat gat tig aaa gat git Val Asn Phe Asp Sor Leu Lys Val Tyr Leu Asp Asp Trp Lya Asp Val 115	384
cuc con can ggt ata agt tog tit act gat gat att cag got ggt aac Lou Pro Gln Gly Ile Ser Sor Pae Ile Asp Asp Ile Gln Ala Gly Asn 130	432
tac ten aca tet tet tet gat gat ett set gat sat tet gen get ggt Tyr Ser Thr Ser Ser Leu Amp Amp Leu Ser Glu Am Pho Ala Val Gly 145 150	480
Lys Gln Leu Leu Arg Asp Tyr Asn Ile Glu Alt Lys Ris Pro Val Val 165	520
and got one ogt one and too and ogt and one ago too one of the Met Val Pro Gly Val Ile Ser The Gly Ile Glu Ser Tre Gly Val Ile	376
God Vab Yab Gor Cha Yab sen Sen You have but ynd Pha ynd Pen Lib der dan dan dan nde Dan ade her dud con pro odt sen rud opd pad	624
gga agt ttt tac atg ctg aga aca atg gtt atg gat aaa get tgt tgg Gly Ser Phe Tyr Met Leu Arg Thr Met Val Met Asp Lys Val Cys Trp 210 215	672
tig aas cat gos atg tta gat cot gos ace got obg gad oca cog aad Lou Lys His Val Net Leu Asp Pro Glu Thr Gly Leu Asp Pro Pro Ann 235 240	720
tot acg ota cgt gca gcz cag ggo the gas toz act gzt tat the atc Phe Thr Leu Arg Ala Ala Gln Gly Phe Glu Ser Thr Asp Tyr Phe Ile 255	765
SCA SES THE USS ALL USS ARE SET THE CAR HAL SUS STA ALL Ala Gly Tym TEP Ile TEP Ash Lys Val Phe Glm Ash Leu Gly Val Ile 250 255	: 51 5
ggo the gam occ and has seg and ago got got gat tog ago of Gly Tym Glu Pro Asn Lys Met Thr Ser Ala Ala Tym Asp Tmp Ang Le 275	± 864 u

Sea tat the gat cha gas aga ego gat agg the tit deg and cha mag 9 Ala Tyr Leu Asp Leu Glu Arg Arg Asp Arg Tyr Phe Thr Lys Leu Lys 290 295 100	12
gas cas are gas etg the cat cas tog age ggt gas ass got of the Glu Glu Glu Ile Glu Leu Phe His Gln Leu Ser Gly Glu Lys Val Cys Leu. 320	
att ggs cat tot atg ggt tot dag att atd tet tad out by: Trp The Gly His Ser Met Gly Sor Gla Ile Ile Phe Tyr Phe Met Ly: Trp 325 325	1008
ged gag ged gas ggd cet obt tac ggt ast ggt ggd dgd gag Val Val Glu Alz Glu Gly Pro Leu Tyr Gly Ann Gly Gly Arg Gly Trp Val 345	1055
Ash Glu His Ile Asp Ser Phe Ile Ash Ala Ala Gly Thr Lou Leu Gly 355 355	1104
get con mag gen get con get eth att agt ggt gan atg ann gat mee Ala Pro Lys Ala Val Pro Ala Leu Ile Ser Gly Glu Met Lya Asp Thr 370 375	1200
att caa tta aat acg tta god atg tat ggt ttg gaa aag ttd tto toa The Gha Leu Asa The Lou Ala Mot Tye Gly Leu Ghu Lys Phe Phe Ser 395 400	1248
AGE ART GAG AGE GIR BAR AND THE CAR AND THE TYP GLY GLY ILE PRO SCY AND ILE GLU AND ALL AND AL	1296
atg cia ccs and ggs gas gag gtc att tgg ggs gat atg nag tca tct Met Leu Pro Lys Gly Glu Glu Val Ile Trp Gly Asp Met Lys Ser Ser 420	1344
toa gag gat goa tog ant dad mad met gad mon tad ggd ant tod mot Ser Glu Asp Ala Leu Asc Asc Asc Thr Asp Thr Tyr Gly Asc Phe Ile 445	
ogs tit gas agg ast acg ago gat got tid asc mas ant tig som and Arg Phe Glu Arg Asn Thr Ser Asp Ala Phe Asn Lys Ash Leu Thr Met 450	
the get got the her and and the tog and the get get the Glu Try hou Glu Lys hap him lie han Met The Lou Ber Ile Ser Pro Glu Try hou Glu 480	

					•														
aga	a	ca	ata		at (CZC	cag	tac	tcg	ttc	585	cac	ttc	229	88	5	25	çaa	1458
270	. 7	- \70	val	н	is	Glu	Gln	Tyr	ser	Phe	Gly	Tyz	Ser	TA	7.5	r. G	Ţu	Glu	
	•					485		-			450					4	95		
									~~~	CAC	885	cac	tes	: LC	: 22	<b>t</b> c	ca	atc	1536
Sag	= 1	E C &	8.≘	i .	.aa	. ZE	222	C 5-4	***	:-	T 1/2	u i e	نتدنل د د د		- As	n P	200	Mez	
Gl	<b>T</b> ]	Г¢л	Ar	<u> </u>	ys.	naA	GIA	Leu	HIE	His	ny.s				. 51	n -			
				5	00				•	EOE						. •			
																. t <b></b>			
57	a	gta	c=	a c	===	CCR	g 2 s.	ācs	ccc	CZC	atg	223	. 22	ב בם		:	LCA	CAC	1584
Gl	u	Val	Pr	o I	Leu	Pro	Glu	Ala	Pro	His	Met	Ly	: Il	e Ty	= C}	/5 ]	TE	тут	
			Ξl						520					<b>52</b>	S				
						٠													
۔۔۔	,,,,,,	<u> ج</u> ہم	2 2		225	cca	act	gaa	. acc	902	. =2:	: gta	i ta	t aa	ğ <u>ğ</u> :	22 9	gag	gat	1632
22	=	2-2	<b>&gt;</b>		»	270	Thr	Glu	Arc	Ala	TV	va.	l Ty	= Ly	5 G	lu (	Glu	Asp	
	У				~3,1	7.0	****	535			•		54	٥					
		530						222											
								-				- ra	r ca	a ao	- a	a=	caa	cet	1680
Gs	10	tcc	===	: C	SCE	crs	aat	בבק	act	. ac	. <u>.</u> .	_ ====	- 61	Sa	- I	vs.	Gla	Pro	
<b>A.</b> 5	7	562	5	EF	VJ=	Leu	Ası	Let	i Thi	- 11	2 ,35	Y	_ G1		- ~	, -		550	
S	15						550	)				55	۵						
													,						
c:	==	tito	: c	LC.	200	gag	999	ga	= 55	a ac	c &r	t cs	g ct	ic gt	:5 5	¢Ğ	CAC	CCA	1728
v:	al	Phe	a L	211	Thr	Glu	Gl	y As	o G1	y Th	r Va	1 P'=	o Le	au Va	11 2	lz	His	Ser	;
,						565				-	57	0					575		
										t gc	÷ 60	a co	:= t:	ac a	ac c	est	gcs	: 655	177 <i>6</i>
£.	55	£\$1	ככ	e.c	25.5			: La		- =-			- T	ه ــِـر	sn I	220	Ala	G13	,
×	et	C.A.	s F.	is	Lys	TI	P AL	9 (1T	E GI	y Al				, — - <u>-</u>		50		_	
					550	)				sa	=							٠	
													_						1624
2	tt	. aa	c ç	==	e. C (	t at	ב בַנ	g ga	a at	.g &8	La Ca	:c c:	sā c	ca g	22 (			_ <u>_</u>	- 1624
I	lc	: As	n V	al	Th	= Il	e Va	.1 G1	u Me	t Ly	/s #.	s C	ln P	TO A	gp.	A.= 9	Pn	e A5)	P
				93					60					E	05				
			. –	-															
_						<u> </u>	<b>.</b>	IA 29	:	:= 5:	12 C	ac g	ta g	ac a	te	ctc	gg	c ag	1872
2		. cg	= =		22	~ 50	. 7.			La G	i ,	is V	al A	so I	le	Leu	. Gl	y Se	<b>=</b>
1	TTC				GI	λ wr	נת ב							20					
		61	.0					6:	15					•			-		
																>		=	.t 1920
Ş	909	9 9 2	5	===	az	c ga	ב ב	1C E	ic t	tg 2	ee e	tt g	<b>=</b> a	igc 5	, <u>-</u> -		. 22	~ 50	
7	41:	a Gl	lu I	Lev	. A.s	n As	j T	yr I	le L	eu ly	ys I	le A	la s	Ser (	31 Y	Ann	GI	~ A:	· <u></u>
	52:				•			30	•			€	35					64	.0
`	<b></b> i	آجم ہم		<u></u>		.,		ar t	to =	st a	22 =	tg a	gc (	cag '	252	gt	: ::	ים כו	ig 1962
	- = -	. 51				-= -	,		-3 - 611 -	בי א		eu S	:= (	Gln '		Val	<u>.</u> 5	= G	in.
	٦e	u.V2	<b>2.</b> ↓	المال	1 P:			للا شقيد	= 4 2		 م	:50			-		G E	5	
						64	15 ~				-	_ ~							

## SUBSTITUTE SHEET (RULE 26)

- 1

1986

ate occ tto coa ate taa Met Pro Phe Pro Met 660

<210> 2 g <211> 651 <212> PRT <213> Saccharomyces cerevisiae <400> 2 Met Gly Thr Lou Phe Arg Arg Asn Val Gln Asn Gln Lys ser Asp Ser Asp Glu Asn Asn Lys Gly Gly Ser Val His Asn Lys Arg Glu Ser Arg 25 Asn His Ile His His Gln Gln Gly Leu Gly His Lys Arg Arg Rrg Gly 40 lic ser Gly Ser Ala Lys Arg Ann Glu Arg Gly Lyn Asp Phe Asp Arg · 55 Lys Arg Asp Gly Asn Gly Arg Lys Arg Trp Arg Asp Ser Arg Arg Leu 75 Ile Phe Ile Leu Gly Ala Phe Leu Gly Val Leu Lau Pro Pho Ser Phe 70 Gly Ala Tyr His Val His Asn Ser Asp Ser Asp Leu Phe Asp Asn Phe 105 Val Ash Pho Asp Ser Lou Lys Val Tyr Leu Asp Asp Trp Lys Asp Val : 120 Leu Pro Gln Gly Ile Ser Ser Phe Ile Asp Asp Ile Gln Ala Gly Asn Tyr Ser Thr Ser Ser Leu Asp Asp Leu Ser Glu Asn Phe Ala Vel Gly Lys Glm Leu Leu Arg Asp Tyr Asm Ile Glu Ala Lys His Pro Val Val 150 Met Val Pro Gly Val Ile Ser Thr Gly Ile Glu Ser Trp Gly Val Ile 185 Gly Asp Asp Glu Cys Asp Ser Ser Ala Mis Phe Arg Dys Arg Deu Tep 200 Gly Ser Phe Tyr Met Leu Arg Thr Met Val Met Asp Lys Val Cys Tip hen Lys His Val Met Leu Asp Pro Glu Thr Gly Lau Asp Pro Pro Asn Pha Thr Leu Arg Ale Ala Gln Gly Pha Glu Sar Thr Asp Tyr Phe Ile 230 250 Ala Gly Tyr Trp Ile Trp Asn Lys Val Phe Gln Asn Leu Gly Val Ile 2 € 5 Gly Tyr Glu Pro Ash Dys Met Thr Ser Ala Ala Tyr Asp Trp Arg Leu 289 . Als Tyr Leu Asp Leu Glu Arg Arg Asp Arg Tyr Phe Thr Lys Leu Lys Glu Gln Ile Glu Leu Phe His Gln beu Ser Gly Glu Lys Val Cys beu 310

Ile Gly His Ser Met Gly Ser Gln Ile Ile Pho Tyr Phe Mcc Lys Trp
Val Glu Ala Glu Gly Pro Leu Tyr Gly Asn Gly Gly Arg Gly Trp Val
340 345 Ala ala Gly Thr Leu Leu Gly
Ash Glu His Ile Asp Ser Phe Ile Ash Ala Ala Gly Thr Leu Leu Gly
Ala Pro Lys Ala Val Pro Ala Leu Ile Ser Gly Glu Met Lys Asp Thr
375  The Gln Leu Asa The Leu Ala Met Tyr Gly Leu Glu Lys Pho Phe Ser  195 400
Arg Ile Glu Arg Val Lys Met Leu Gln Thr Trp Gly Gly Ile Pro Sar
Met Leu Pro Lys Gly Glu Glu Val Ile Trp Gly Asp Met Lys Ser Ser
) 420 Ser Glu Asp Ala Leu Ash Ash Ash Thr Asp Thr Tyr Gly Ash Phe Ile
435 Arg Pho Glu Arg Asn Thr Ser Asp Ala Phe Asn Lys Asn Leu Thr Ket 460
450  455  Lys Asp Ala Ile Asa Met Thr Leu Ser Ile Ser Pro Glu Trp Leu Gla 480
Aff Arg Val His Glu Gln Tyr Ser Phe Gly Tyr Ser Lys Asn Glu Glu Arg Arg Val His Glu Gln Tyr Ser Phe Gly Tyr Ser Lys Asn Glu Glu 495
485 Glu Lou Arg Lys Asn Glu Leu His His Lys His Trp Ser Asn Pro Mct 510
Glu Val Pro Lou Pro Glu Ala Pro Hic Met Lys Ile Tyr Cys Ile Tyr
Clu Val Aca Aca Pro The Glu Are Ala Tyr Val Tyr Dys Gir
And Sen Sen Ald Den Ash Den The Ile And Tyr Glu Sen Dyn Glu Pro
545  Val Phe Leu Thr Glu Gly Asp Gly Thr Val Pro Leu Val Ala His Ser  575
565  Met Cys His Lys Trp Ala Gln Gly Ala Ser Pro Tyr Asn Pro Ala Gly  590
The Mai Glu Mat Lys Mis Gin pro Asp Ang sat Ang
. <b>E</b> ON
Ile Arg Gly Gly Ala Lys Ser Ala Glu His Val Asp Ile Leu Gly Ser
Ala Glu Leu Asn Asp Tyr Ile Leu Lys Ile Ala Ser Gly Asn Gly Asp
625 630 Leu Val Glu Pro Arg Gln Leu Ser Asn Leu Ser Gln Trp Val Ser Gln 655
645 645 650
0 9 3
Met Pro Phe Pro Mat
650

<210> 3 &

<211> 312 <211> 2312 <212> genom <213> Schiz <400> 3	ic DNA osaccharomy	edmoq. sec			
ATEGCGTCTT	CCAAGAAGAG	CAAAACTCAT	ARGARRAGA	AAGAAGTCAA	<b>5</b> 0
ATCTCCTATC	GACTTACCAA	ATTÇAAAGAA	ACCARCTOSC	GCTTTGAGTG	100
AGCARCCTIC	AGCGTCCGAA.	ACACAATCTG	TTTCAAATAA	ATCAAGAAAA.	150
TCTAAATTTG	GRAAAAGATT	GAATTTTATA	TTGGGCGCTA	TTTTGGGAAT	200
ATGCGGTGCT	TTTTTTTTCG	CTGTTCGAGA	CGACAATGCT	GTTTTCGACC	250
CTGCTACGTT	AGATAAATTT	CGGAATATGC	trecestate	AGACTTGTTT	300
GATCACATTA	AREGRIATIT	ATCTTATAAT	GTGTTTAAGG	ATGCACCTTT	350
TACTACGGAC	AAGCCTTCGC	AGTCTCCTAG	CGGARATGAR	GTTCAAGTTG	400
G'CCTTGATA.T	GTACAATGAĞ	GGATATCGAA	GTGACCATCC	TGTTATTATG	450
GTTCCTGGTG	TTATCAGCTC	AGGATTAGAA	AGTIGGICGI	TTAATAATTG	500
CTCGATTCCT	TACTTTAGGA	AACGTCTTTG	GGGTAGCTGG	TCTATGCTGA	550
AGGCAATGTT	CCTTGACAAG	CARTGCTGGC	TTGAACÁTTT	AATGCTTGAT	€00
AAAAAAACCG	GCTTGGATCC	GAAGGGAATT	AAGCTGCGAG	CAGCTCAGGG	650
GTTTGARGCA	GCTGATTTTT	TTATCACGGG	CTATTGGATT	TGGAGTAAAG	700
TAATTGAAAA	CCTTGCTGCA	ATTGGTTATG	AGCCTAATAA	CATGTTAAGT	750
GCTTCTTACG	ÄTTGGCGGTT	ATCATATGCA	AATTTAGAGG	AACGTGATAA	200
AINTTITTCA	ARGITAAAA.	TGTTCATTGA	GTACAGCAAC	ATTGTACATA	850
AGAAAAAGGT	AGTGTTGATT	TCTCACTCCA	TGGGTTCACA	GGTTACGTAC	900
· TATTTTTTTA	AGTGGGTTGA	AGCIGAGGGC	TACGGAAATG	CADOCADID	950
TIGGGTTAAT	GATCATATIG	AAGCATTTAT	ARATGTEAGT	CTCGATGGTT	1000
GTTTGACTAC	GTTTCTAACT	TTTGAATAGA	TATOGGGATO	TTTGATTGGA	1050
GCAGCGAAAA	CAGTGGCAGC	GCTTTTATCG	GGTGAAATGA	AAGATACAGG	1100
TATIGTAATT	ACATTAAACA	TGTTAATATT	TARTTTTTGC	TAACCGTTTT	1150
ANGCTCARTT	GAATCAGITI	TOGGTCTATE	GGTAAGCAAT	AAATTGTTGA	1200
GATTTUTTAC	TANTITACIG	TITAGTTEG	TTTTTTAAAAA	TICCCCTICI	3250
GASSTATATT	CAAAAATACA	AATGTGCTCT	ACTITITEES.	. ACTTITAATA	1300
GAGAGCCATG	: ATGGTTCGCA	CTATGGGAGG	AGTTAGTTCT	AIGCTTCCTA	1350
AAGGAGGCGA	. TGTTGTATGG	GGAAATGCCA	. GTTGGGTAAC	OPTETATAA :	1400

TOLKSCAAN	1450
TGTTAATTT TTATTAATAT TTAGGCTCCA GATGATCTTA ATCAAACAAN	1500
TTTTTCCAAT GGTGCAATTA TTCGATATAG AGAAGACATT GATAAGGACC	1100
ACGATGARTT TGACATAGAT GATGCATTAC AATTITTAAA AAATGTTACA	1550
GATGACGATT TTAAAGTCAT GCTAGCGAAA AATTATTCCC ACGGTCTTGC	1600
TIGGACIGAA AAAGAAGIGI TAAAAAATAA CGAAAIGCCG TCTAAAIGGA	3,650
TAMATOCGOT AGAAGTAAGA ACATTAAAGT TACTAAATTA TACTAACCCA	1700
NATAGACTAG TOTTCOTTAT GOTCOTGATA TGANAATTTA TIGOGTTCAC	1750
GGGGTCGGAA AACCAACTGA GAGAGGTTAT TATTATACTA ATAATCCTGA	1900
GGGGCAACCT GTCATTGATT CCTCGGTTAA TGATGGAACA AAAGTTGAAA	1950
ATGTGAGAGA ATTTATGTTT CAAACATTCT ATTAACTGTT TTATTAGGGT	1900
ATTOTTATEG ATGATEGTEA TEGAACTTTA CCAATATTAG CCCTTEGTTT	1950
GGTGTGCAAT RAAGTTTGGC AAACAAAAAG GTTTAATCGT GCTAATACAA	2000
GTATCACAAA TTATGAAATC AAGCATGAAC CTGCTGCGTT TGATCTGAGA	2050
GGNGGACCTC GCTCGGCAGA ACACGTCGAT ATACTTGGAC ATTCAGAGCT	2100
AARTGTATGT TCATTTTACC TTACAARTTT CTATTACTAA CTCTTGAAAT	2150
AAGGAMATIA TITTAAAAGT TICATCAGGC CATGGTGACT CGGTACCAAA	2300
COGTTATATA TOAGATATOO AGTACGGACA TARGTTTTGT AGATTGCAAT	2250
CCGTTATATA TCAGATATCO AGIACOCOTA AGIAGATARA TCTCGATARA TALCACTARA ACCGARACAGO GARATARTAR ATGAGATARA TCTCGATARA	2300
TAACTAACTA ACCGAACAGG GAARLAA	2312
CCTAGAAATT AA	

ATGCCCCTTA TTCATCGGAA AAAGCCGACG GAGAAACCAT CGACGCCGCC 50 ATCTGAAGAG GTGGTGCACG ATGAGGATTC GCAAAAGAAA CCACACGAAT 100 CTTCCARATO COACCATARG ARRICGARCG GROGROGGRA GTGGTCGTGC 150 200 ATCGATTCTT GTTGTTGGTT CATTGGGTGT GTGTGTAA CCTGGTGGTT TUTTOTOTTO CTTTACAACG CAATGCCTGC GAGCTTCCCT CAGTATGTAA 250 CGGAGCGNAT CACGGGTCCT TTGCCTGACC CGCCCGGTGT TAAGCTCAAL 300 AAAGAAGGTC TTAAGGCOAA ACATCCTGTT GTCTTCATTC CTGGGATTGT 350 CACCGOTGGG CTCGAGCTTT GGGAAGGCAA ACAATGCGCT GATGGTTTAT 400 TTAGAAAACG TTTGTGGCGT GGAACTTTTG GTGAAGTCTA CAAAAGGTGA 450 500 CCTCAACAAT TCTCACTCTT CCTTTATATT GGGATTTGGA TTGGATCTGA TGAGATCACG CACTTGTTGC TTCTTCAACA TCACTCAAAC TTTAATTCCA 550 TGTTTGTCTG TCTTACTCTT TACTTTTTTT TTTTTTTGAT GTGAAACGCT 600 ATTITCTTAA GAGACTATTT CTGTATGTGT AAGGTAAGCG TTCCAAGGAC 650 GTARTGGGT TEGRCTATTT CTGTTTGRTT GTTARCTTTR GGATATARA 700 TAGCTGCCTT GGAATTTCAA GTCATCTTAT TGCCAAATCT GTTGCTAGAC 750 ATGCCCTAGA GTCCGTTCAT AACAAGTTAC TTCCTTTACT GTCGTTGCGT 200 GTAGATTTAG CTTTGTGTAG CGTATAATGA AGTAGTGTTT TATGTTTTGT 250 TGGGARTAGA GAAGTTCTAA CTACATCTGT GGARAGTGTG TTCAGGCTGT 900 GATAGAGGAC TGTTGCTTTA TTATTCARCT ATGTATATGT GTAATTAAAG 950 CTAGTTCCTT TTTGATCTTT CAGCTCAATG TGCTTTTCTC AATITTTTTC 1000 TONATTICAA AGITTCACAT CGAGTTTATT CACATGTCTT GAATTTCGTC 1050 CATCCTCGTT CTGTTATCCA GCTTTGAACT CCTCCCGACC CTGCTATGGA 1100 TATNITARAR ARRAGIGIT TIGTGGGTTG CATCHIGTT ACGATCIGCA 1150 TOTTOTTOTT TOGGCTOAGT GTTCATGTTT TTGCTATGGT AGAGATGGGC 1200 ARTOTTATIG TIGHTGGTAR CAGTGGTATA GTTGATAGTA TOTTARCTAR 1250 TOANTATOT CITTONITCH GGCCTCTATG TIGGGTGGAA CACATGTCAC 1300 TTGACAATGA AACTGGGTTG GATCCAGCTG GTATTAGAGT TCGAGCTGTA 1350

TCAGGACTCG TGGCTGCTGA CTACTTTGCT CCTGGCTACT TTGTCTGGGC	1400.
AGTGCTGATT CCTAACCTIG CACATATIGG ATATGAAGAG AAAAATATGT	1450
ACATGGCTGC ATATGACTGG CGGCTTTCGT TTCAGAACAC AGAGGTTCTT	1500
TTCTCATCGT TCTTTCTATT ATTCTGTTCC ATGTTACGTT TCTTTCTTCA	1550
TTCTCATCGT TCTTTCTALL ALLCCATCATCA ARTTAATAGG TACGTGATCA	2600
TTACTTAAGG CTTAAATATG TTTCATGTTG AATTAATAGG TACGTGATCA	1,650
GACTOTTAGO CGTATGAAAA GTAATATAGA GTTGATGGTT TCTACCAACG GTGGAAAAA AGCAGTTATA GTTCCGCATT CCATGGGGGT CTTGTATTT	1700
GTGGARARA AGCAGTTATA GIICCGCCT	1750
CTACATTITA IGAAGIGGGT IGAGGCACCA GCTCCTCIGG GIGGCGGGG	1900
TGGGCCAGAT TGGTGTGCAA AGTATATTAA GGCGGTGATG AACATTGGTG	1850
GACCATITCT TGGTGTTCCA AAAGCTGTTG CAGGGCTTTT CTCTGCTGAA	1900
CONNEGATO TICCAGTIGO CAGGIATIGA ATATOIGCIT ATACITICA	
COLTENED OF CITEGOTOTE GARCTCARAG TTATTCTACT AGAINTCAL	1950
TCTANTANCA TIGCTATATT ATCGCTGCAR CTGACATTGG TTGATTATTT	2000
TTGCTGCTTA TGTAACTGAA ACTCTCTTGA GATTAGACAA ATGATGAATT	2050
TTGCTGCTTA TGTAACIGAA ACITTOTTA	2100
GATAATTCTT ACGCATTGCT CIGIGATCH	2150
TAXCATTIGE CATACTGEET TITGGAGGGC ATTGAATETT GCTATGGAAA	2200
GCGCTGGAGC TTCCATGCTT GCATTCTTTA CCAATTAGCG TTATTCTGCT	2250
TOTTCAATT TTOTTGTATA TGCATCTATG GTCTTTATT TCTTCTTAAT	2300
TARAGACTEG TIGGATTAGT IGCTCTATTA GTCACTIGGT ICCTTAATAT	2350
ACARCTITAC TITCTICGAA NATIGCAGAG CGATIGCCCC AGGAILCIAN	•
GACACCGATA TATTTAGACT TCAGACCTTG CAGCATGTAA TGAGAATGAC	2400
ACGCACATGG GACTCAACAA TGTCTATGTT ACCGAAGGGA GGTGACACGA	2450
TATEGGGCGG GCTTGATTGG TCACCGGAGA AAGGCCACAC CTGTTGTGGG	2500
TATEGGGCGG GCTTGATIGG TCATGGGT GAAGCAGGTG AAAACGGAGT  NAAAAGCAAA AGAACAACGA AACTTGTGGT GAAGCAGGTG AAAACGGAGT	2550
NARAGERAA AGARCAACGA MADITETOTO	2600
TTCCAAGAAA AGTCCTGTTA ACTATGGAAG GATGATATCT TTTGGGAAAG	2550
ANGTAGCAGA GGCTGCGCCA TCTGAGATTA ATAATATTGA TTTTCGAGTA	2700
ACGRETATA ARTCATATA ARCETTGTAC ATTTTGTGAT TGTRTGATGA	2750
AEGAERIRIA ARTONIPOTO GEGRAGGGEG CEGECARAGG ECAGAGEREC ATRECEGRA ATTERCEGE GEGRAGGGEG CEGECARAGG ECAGAGEREC	2300
CONSTRUCTED CONTROL CO	
TGCTGGGATC AAAGCTATCG CTGAGTATAA GGTCTACACT GCTGGTGAAG	2830

## SUBSTITUTE SHEET (RULE 26)

CTATAGATET ACTAGATIAT GTTGCTCCTA AGATGATGGC GCGTGGTGCC	2900
CTATAGATET ACTAGATIAT GITOTTATAGATGAT TTGGATGACA CCAAGTATCA	2950
GCTCATTTCT CTTATGGAAT TGCTGATGAT	3000
AGATOCCAAA TACTGGTCAA ATCCGTTAGA GACAAAGTAA GTGATTTCTT	3050
GATTCCAACT GTATCCTTCG TCCTGATGCA TTATCAGTCT TTTTGTTTTC	3100
GGTCTTGTTG GATATGGTTT TCAGCTCAAA GCTTACAAAG CTGTTTCTGA	3150
GCCTTICTCA AAAAGGCTIG CTCAGTAATA TIGAGGIGCT AAAGTIGATA	3200
TOTAL ATCCTCCGTT TGGTTTGTTC TGGTTTTCA	3250
TOTTCCGAR TOTTCCTGAG ATGGAAATCT ACTCATTATA CGGAGIGGGA	3300
STIRTINGE ANGRAGEATA CGTATACAAG CTTAACCAGT CICCOGAGA	
THE TEST TITCAGNIAT TOACTICIGO TOACGAGGAG GAUGAAGAIA	3350
POTENTIAL DECARGAGIT TACARTETES ATGESCATER ANCASTROLL	3400
GTCCTAAGTG CCGGGTACAT GTGTGCAAAA GCGTGGCGTG	3450
ATTCHACCOT TOOGGAATCH AGACTTATHT AAGAGAATHC AATCACTOTO	350,0
ATTCAACCCT TECGGARICATOR TO THE COCCATGOR TECCCATGOR TECCCATGOR TECCCATGOR TECCCATGOR TECCCATGOR TECCCATGOR TECCCATGOR TECCCATGOR TECCATGOR TECCATGOR TECCATGOR TECCATGOR TECCATGOR TECCATGOR TECCATGOR TECCATGOR TECCATGOR	3550
CGCCGGCTAR CCTGTTGGRA GCCTTGCCGC GATATCATGA GGGTTGCCGC	3600
GATATONIGG GARACTITGO IIIGATONIA COAGGICCAC TOIGGOATAT CGGAGGIAAC GGGICIGATA TAGGACAIGA CCAGGICCAC TOIGGCATAI	3550
CGGAGGTAAC GGGTCTGATA TACGACATON	3625

<210> 8 \( \text{P} \)
<211> 616
<211> CDNA
<213> Neurospora crassa
<400> 8

GGTGGGGAAG ACGANGGGGG AAG	TTGGAGG	CTARCGAGAA	TGACNCTCGG	•
AGATGGATCT ACCCTCTAGA GAC				ioo
GINTACNGTT INTATGGGTA GGA				150
TEGCGCCCGA TCCCGGGACG ACI				200
ACTITGACTH AGGGGCACAT TG				250
TGGCACAGTG AACCTTATGA GT				300
ARATGRAGAG ATACRATCCT GC				350
CCGCAIGAAC CAGAACGGTT CA				500
CTTAAATATG TAGAAAAGGT TG				SSO
ACATAGGTTA CTCAATAGTA TG				600
				616
AAAAAA AALLIKKKK				

PCT	ÆP00	0/02701
-----	------	---------

50	٥/	53	
----	----	----	--

WO 00/60095

TTANCAGAGG	AGCTGATGAC	TCCGGGAGTG	CCAGTCACTT	GCATATATGG	1300
GAGAGGAGTT	GATACACCGG	AGGTTTTGAT	GTATGGAAAA	GGAGGATTCG	1350
ATAAGCAACC	AGAGATTAAG	TATGGAGATG	GAGATGGGAC	GGTTAATTTG	1400
GCGAGCTTAG	CAGCTTTGAA	AGTCGATAGC	TTGAACACCG	TAGAGATTGA	1450
TGGAGTTTCG	CATACATOTA	TACTTAAAGA	CGAGATCGCA	CTTAAAGAGA	1500
TTATGAAGCA	GATITCAATT	ATTAATTATG	AATTAGCCAA	TGTTAATGCC	1550
GTCAATGAAT	GA	•	•		
					1552

<210> 9 \( \text{V} \)
<211> 1562
<212> genomic DNA
<213> Arabidopsis thaliana
<400> 9

				CTCDTAGCGA	TACTCGTTGT	50
					TACTCGTTGT,	
					GTGTACCCTT	100
,	TGATTCTGGT	TCCAGGAAAC	GGAGGTAACC	AGCTAGAGGT	ACGGCTGGAC	150
					TATATCCGAT	200
,	DAADAATAJI	AGTGGTGGAT	GGTTTAGGCT	ATGGTTCGAT	GCAGCAGTGT	250
	TATTGTCTCC	CTTCACCAGG	TGCTTCAGCG	ATCGAATGAT	GTTGTACTAT	00É
					AAACCCGGGT	350
	TCCTCATTTC	GGTTCGACCA	AATCACTTCT	ATACCTCGAC	CCTCGTCTCC	400
	GGTTAGTACT	TTCCAAGATA	TATCATTTTG	GGACATTTGC	ATAATGAACA	450
	AAATAGACAT	AAATTTGGGG	GATTATTGTT	ATATCAATAT	CCATTTATAT	500
					G TGAGTGTTAT	550
					ANTANTGTIG	600
	CTATGTCATC	AGAATTATA	A GGACACTATO	TARATGTAG	TTAATAATAA	650
	GGTTTGATTI	r GCAGAGATG	C CACATOTTA	ATGGAACAT	r tegtgllige	700
)					CTAGGAGCIC	750
					C CTCCCGTGTA	800.
					A AAACTAGCAG	850
	CG)GADCGA	z GGAAAGCCA	G TGATACTCC	T CTCCCATAG	CCTAGGAGGAC	900
	<b>中でででいる。</b>	T CCATTICCI	C AACCGTACC	A CCCCTTCAT	G GCGCCGCAAG	950
					G GGACGATCTC	1000
					C CCTTTAGTTA	1050
	TCAGATGAA	e Mewities	c circlgles	a cotocgaga	G TAACCAATGG	1100
	ACCCTTIGC	T GUTCAGALL		c agaactaaa	C CCCTTCTCGT	1150
	CINCTICCA	T CIACCAAA	al Gilleway			1200
	AACTCCCCA	G GTTAACTA	IR CAGITTACE	A CATGGAICG	R TTTTTTGCAG	
	ACATTGGAT	T CTCACAAG	er ettetect	T ACAAGACAA	G AGTGTTGCCT	1250

<210> 10 1/V

```
<211> 3596
 <212> genomic DNA
 <211> Arabidopsis thaliana
 c400> 10
 ATGGGAGGGA ATTCGAAATC AGTAACGGCT TCCTTCACCG TCATCGCCGT
 TTTTTCTTG ATTTGCGGTG GCCGAACTGC GGTGGAGGAT GAGACCGAGT
 TICACGGCGA CTACICGAAG CTATCGGSTA TAATCATTCC CCCAFTTCGS
                                                             150
 TCGACGCAGC TACGAGCGTG GTCGATCCTT GACTGTCCAT ACACTCCGTT
                                                             200
 CGACTTCAAT CCGCTCGACC TCGTATGGCT AGACACCACT AAGGTCCGTG
                                                             250
 ATCTTCATTT CCTTCGCTCC TTATTCTGTC GGTCGAGTCA CTTGTTGATG
 TCGTTCATTA GTCAACAGTG ACGCTTCTGA ATCTGAGTTT AGAGTCATAT
 AAAACAGCTO ACTCGGCGAG TOTTTCCCAT CGCTTTTGGT TCGCTAAATG
                                                             450
 TAGCGCANTG AATGTGTAAT TAGTCTGCGC TTTTTATTCA ACTAGATCTG
                                                             500
CANGITTITC AGAGIGCTCA ATAGTAGTTA GARAATGTTA GGTCATTTTA
                                                             550
 CTTGTGCATT GTGATTCTTT TGGTTGTTGC TTACTGATCG ACGTGATGGA
                                                             600
                                                             650
 TGGTTTACAG CTTCTTTCTG CTGTCAACTG CTGGTTTAAG TGTATGGTGC
 TAGATOCTTA TAATCAARCA GACCATOOOG AGTGTAAGIC ACGGCCTGAC
                                                             700
 AGTGGTCITT CAGCCATCAC AGAATTGGAT CCAGGTTACA TAACAGGTAG
                                                             750
 TTTCGGATTT TTCTTTCTTT TGAGTTTTCT TCAATTTGAT ATCATCTTGT
                                                             800
 TGTGATATA: TATGGCTANG TTCATTAATT TGGTCAATTT TCAGGTCCTC
 TITCTACTGI CTGGAAAGAG TGGCTTAAGI GGTGTCTTGA GTTTGGTATA
                                                             900
 GARCCANNIG CARTIGICGC IGTICCATAC GRITGGAGAT IGTCACCANC
                                                             950
 CARATTGGAA GAGCGTGACC TTTACTTTCA CAAGCTCAAG TTAGTCCTTA
 TOAGGCTART GTCTTTTATC TTCTCTTTTT ATGTAAGATA AGCTAAGAGC
                                                             1050
                                                            1100
 TOTGGTCGTC TICCTITTIG CAGGTTGACC TITGAAACTG CTTTAAAACT
 CCGIGGCGGC CCTTCTATAG TATTTGCCCA TTCAATGGGT AATAATGTCT
 TONGREACTY TOTGGARTGG CTGRGGCTAG ARATTGCRCC RARRORTRAT
                                                             1250
 TTGRRGTGGC TTGRTCAGCA TATCCATGCT TATTTCGCTG TTGGTACCGG
 CCTACTATCC TTANGTTACC ATTTTATTTT TTCTCTAATT GGGGGAGTTA
 TGTTGTGACT TACTGGATTG AGCTCGATAC CTGATTTGTT GTTGATTTAG
                                                             1350
 GAGCTECTET TETTGGTTET GTTGAGGCAP. TEAAATCTAE TETETETGGT
                                                             1400
 GTAACGTTTG GCCTTCCTGT TTCTGAGGTG ACCTCTGACT TCTCTTTAGT
                                                            1450
 TITALGITAGI TGATATCARC CAGGICTIAT ARCTCACIGG ATTITCCITI
                                                             1500
  TGARAGTATT ACTITIGITA ATTGARCTGC TGTACGCGAT ATGGTATCTG
                                                             1550
  TAGATCTIGA AGTGCTAGTT ATCAAAGAAC ATATTGTGGG TAGTATACCT
                                                             1600
 GTCAGGGGCC TTAGGTAATA CAACCAAACC ACATGTACAC TGATTTAGTT
                                                             1650
 TTCAGATTAT TATGGTAGAC TTTAAGTTGA CAAGAAACTT TGACTGAAAT
                                                             1700
  CITTITATIT TARTAGGCTA TGATTIGTIT ATTGARATCA TGTGACATAT
                                                            1750
  TGACATGCGC TTCTCATGTT TTTTGTTGGC AAGGCTTCAG GGAACTGCTC
                                                           1600
  GGTTGTTGTC CAATTCTTTT GCGTCGTCAT TGTGGCTTAT GCCATTTTCA
                                                             1250
  AAGAATTGCA AGGGTGATAA CNCATTCTGG ACGCATTTTT CTGGGGGTGC
                                                             1900
  TGCMAGAAA GATAAGCGCG TATACCACTG TGATGAAGAG GAATATCAAT
                                                             1950
  CAAAATATTC TGGCTGGCCG ACAAATATTA TTAACATTGA AATTCCTTCC.
                                                              200
  ACTAGOGGTT AGACTOTGTA TATGOAACTG TAACACTAAC AAAAGTTTCA
                                                             2050
                                                             2100
  CCAAGAATGT TCACTCTCAT ATTTCGTECC TTTGATGTGT ATCCATCAGT
                                                             2150
  TACAGAAACA GCTCTAGTCA ACATGACCAG CATGGAATGT GGCCTTCCCA
  CCCTTTTGTC TTTCACAGCC CGTGAACTAG CAGATGGGAC TCTTTTCAAA
  GENATAGAAG ACTATGACCC AGATAGCAAG AGGNTGTTAC ACCAGTTAAA
                                                             2300
  GAAGTACGTA CCTTTCTTTG TGATAAGAAA TATTGCTCAT CGATCATCAC
  TIGGIGGCIT CITGIACGIC AAATIGITIT GITTAAATCI CIATAICAAT
                                                             2350
  TGTTCATATG CTTTGTCTTT CTTACTATAA GAAACAAGTA TAATCAGAAA
                                                             2400
  CCTTATTATT GATTATCAGT TCTCTCCTTA TATTATGGAA TGTCTTTTTC
                                                             2450
 GUITACAGIT ATGANIGCAR ARGGGGGTAT TITAGITGAT IGATICICIC
                                                             2500
  ATTCTCTAGT TTGTTTTGAC TAATAGCGTC BATTTTGTTT TTCTAGCAAA
                                                             2550.
  TOTTTGTGAN, TTATATATAN CATGOTRACT ATACTTTTCA GGTTGTATCA
  TGATGACCCT GTTTTTAATC CTCTGACTCC TTGGGAGAGA CCACCTATAA .
  ANATOTATT TIGCATATAT GGTGCTCATC TAAAGACAGA GGTATGATGC
  ATTCTCAATA TCACATTATG CGTTGACTTT GTTATTATAT TCCCCATTTG
```

TOTTCTCCT TOCATCTTAT	2800
	2850
GTTTGCAATA TCTTTTTGAA TIATGTAT GAAGCTGTCT GTCATAGGTT GCTATTAAGC GTTAAAGGTA CTAAATGTAT GAAGCTGTCATA ATTGGATCAT	2900
	2950
CGTTALLACL SIGGEST CCTCGTGTCA ACCURATE	3000
CACGGALATE ALLEGE AAGTCTICLE INTERPRETATION OF THE PROPERTY OF	3050
CCCCAATGGC ACCAAA TTATTAAACA ACCAAA TTATTAAACA	3100
GTGGCAIGIT ALTERNATION AGTGGATGAL CALL	3150
AAGIAGETE TO THE TOTAL CACILGIAN TO THE TOTAL CACILGIAN	3200
AAGIGGAAG AGGIATIAL	3250
ACCAAACAA AACTAA	3300
CIGCIFFIAN ANNIES - CONTROL TO TGGTIGATGO	3350
GATTGTGCAA TATCTGCACG TOTAL GAAGTTGGTF TIGAACTTA	3400
CCTATARCIO GUARIANTE TACTIGUIA	3450
CITCLIGEAA ACIACIA TAT TOACIALLO CITCLIGUE	3500
CTTGCTA1G1 1C1CA.	3550
TGATIATOAA ALLONDON TARCAATOGO	3500
	3650
CTCTTTTTA GALLES TO TO THE TOTAL TO THE TOTAL TO	3700
	3750
	3800
GAAATAINA CACATAGETA TOATAGETA	3850
	3896
GCACCAAGGG TTAAGTACAT AACCTTTTAT GAAAAGTGGG TATTAA	

<210> 11D <221> 709 <212> CDNA <211> COMACO <400> 11

		٦				
(	CTGGGGGCCAA	AAGTGAACAT	AACAAGGACA	CCACAGTCAG	AGCATGATGT	. 50
•	CAGATOTAC	AAGTGCATCT	aaatatagag	CATCAACATG	GTGAAGATAT.	joo
(	CATTCCCAAT	ATGACAAAGT	TACCTACAAT	GAAGTACATA	açctattatg	150
1	AGGATTCTCA	AAGTTTTCCA	GGGACAAGAA	CAGCAGITTG	GGAGCTTGAT	200
	NAGCAANTC	ACAGGAACAT	TGTCAGATCT	CCAGCTTTGA	TGCGGGAGCT	250
)	GTGGCTTGAG	ATGTGGCATG	ATATTCATCC	TGATAAAAAG	TCCAAGTTTG	300
,	TTACAAAAGG	TGGTGTCTGA	TCCTCACTAT	TTTCTTCTAT	AAATGTTTGA	350
	GTITGTATTG	ACATTGTAAS	TATTGCAACA	AAAAGCAAAG	CCTGGGCCTC	400
	TGAGGGATGA	GGACTGCTAT	TGGGATTACG	GGAAAGCTCG	ATGTGCATGG	450
	GCTGAACATT	GTGAATACAG	GTTAGAATAT	TCARATTATA	TTTTGCAAAA	500
	TATTCTCTTT	TIGIGTATTT	AGGCCACCTT	TCCCCGGTCA	CAACGATGCA	550
	GATATGTATT	CGGGGATGTT	CACCTGGGAC	AGAGTTGCAG	ATTGAAGAGT	. 600
	TCTACAICTC	ACATCCTGTC	ACACTATGTG	TGATATTTAA	GAAACTTTGT	650
	TTGGCGGAAC	<i>NACAAGTTTG</i>	CACAAACATT	TGAAGAAGAA	AGCGAAATGA	700
	TICAGAGAG			•		709

### (12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

## (19) World Intellectual Property Organization International Bureau





### (43) International Publication Date 12 October 2000 (12.10.2000)

PCT

## (10) International Publication Number WO 00/60095 A3

- (51) International Patent Classification⁷: C12N 15/54, 9/10, 15/81, 15/82, 1/16, 5/10, A01K 67/027, C12P 7/64
- (21) International Application Number: PCT/EP00/02701
- (22) International Filing Date: 28 March 2000 (28.03.2000)
- (25) Filing Language:

English

(26) Publication Language:

English

- (30) Priority Data:
  99106656.4 1 April 1999 (01.04.1999) EP
  99111321.8 10 June 1999 (10.06.1999) EP
  60/180,687 7 February 2000 (07.02.2000) US
- (71) Applicant (for all designated States except US): BASF PLANT SCIENCE GMBH [DE/DE]; D-67056 Ludwigshafen (DE).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): DAHLQVIST, Anders [SE/SE]; Hemmansvägen 2, S-244 66 Furulund (SE). STAHL, Ulf [SE/SE]; Liljegatan 7b, S-753 24 Uppsala (SE). LENMAN, Marit [SE/SE]; Revingegatan 13a, S-223 59 Lund (SE). BANAS, Antoni [SE/PL];

Wiolinowa 14, PL-08110 Siedlee (PL). RONNE, Hans [SE/SE]; Dirigentvägen 169, S-756 54 Uppsala (SE). STYMNE, Sten [SE/SE]; Torrlösa 1380, S-269 90 Svalöv (SE).

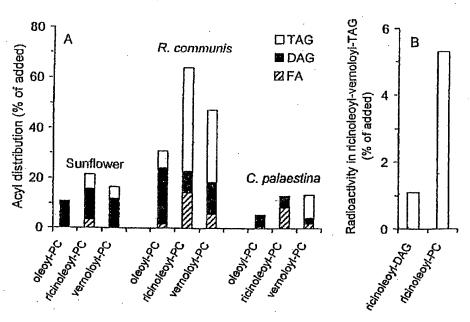
- (74) Agent: FITZNER, Uwe; Lintorfer Str. 10, D-40878 Ratingen (DE).
- (81) Designated States (national): AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

#### Published:

With international search report.

[Continued on next page]

(54) Title: ENZYMES OF THE BIOSYNTHETIC PATHWAY FOR THE PRODUCTION OF TRIACYLGLYCEROL AND RE-COMBINANT DNA MOLECULES ENCODING THESE ENZYMES



(57) Abstract: The present invention relates to the isolation, identification and characterization of nucleotide sequences encoding an enzyme catalysing the transfer of fatty acids from phospholipids to diacylglycerol in the biosynthetic pathway for the production of triacylglycerol, to the said enzymes and a process for the production of triacylglycerols.

N/60095 A

## WO 00/60095 A3



(88) Date of publication of the international search report: 1 February 2001

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

al Application No PCT/EP 00/02701

A. CLASSIFICATION OF SUBJECT MATTER IPC 7 C12N15/54 C12N9/10

C12N5/10

A01K67/027

C12N15/81 C12P7/64

C12N15/82

C12N1/16

According to International Patent Classification (IPC) or to both national classification and IPC

### B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols): IPC 7 C12N A01K C12P

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

STRAND, EPO-Internal, WPI Data, MEDLINE, CHEM ABS Data, BIOSIS, EMBL

Category *	Citation of document, with indication, where appropriate, of	the relevant passages	Relevant to claim No
	PETER VERHASSELT ET AL.: "Tw reading frames revealed in th segent flanking the centromer Saccharomyces cerevisiae chro	e 23.6kb e on the	1-23,27
	right arm" YEAST, vol. 10, no. 7, July 1994 (19 1355-1361, XP002112572	94-07), pages	
	abstract; table 2 -& Swissprot Database Entry Y Accession number P40345; 1 Fe		1-23,27
	XP002112574 the whole document		
	·	-/	
		•	
χ Funt	her documents are listed in the continuation of box C.	χ Patent lamily members are liste	ed in annex.
Special ca	tegories of cited documents :	"T" later document published after the in	nternational filing date
consid	ent defining the general state of the art which is not lered to be of particular relevance	or priority date and not in conflict w cited to understand the principle or invention	th the application but theory underlying the
filing d		"X" document of particular relevance; the cannot be considered novel or cannot be considered novel or the	not be considered to
which	cument which may throw doubts on priority claim(s) or involve an inventive step when the document is thick is cited to establish the publication date of another addition or other special reason (as specified) "Y" document of particular relevance; the claimed inventive step when the document is the document of particular relevance; the claimed inventive at the considered to involve an inventive step when the document is the document is the document is the document is the document involve an inventive at the document is the document is the document in the document is the document in the document is the document involve an inventive at the document is the document in the document in the document is the document in the document		claimed invention
other n	ent referring to an oral disclosure, use, exhibition or neans ent published prior to the international filing date but	document is combined with one or a ments, such combination being obv in the art.	nore other such docu- ious to a person skilled
later th	nan the priority date claimed	. "&" document member of the same pate	<del></del>
ate of the a	actual completion of the international search	Date of mailing of the International s	search report
. 1	7 October 2000	30/10/2000	
ame and 17	nailing address of the ISA  European Patent Office, P.B. 5818 Patentlaan 2  NL - 2280 HV Rijswijk  Tol. (-2,1,70) 340, 2000, Tx, 31,651,ccc.cl	Authorized officer	
	NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Montero Lopez, B	

al Application No PCT/EP 00/02701

a. classification of subject matter IPC 7 C12N15/54 C12N9/10

C12N5/10

A01K67/027

C12N15/81 C12P7/64

C12N15/82

C12N1/16

According to International Patent Classification (IPC) or to both national classification and IPC

#### **B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

C12N AO1K C12P IPC 7

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

STRAND, EPO-Internal, WPI Data, MEDLINE, CHEM ABS Data, BIOSIS, EMBL

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to daim No.
X	PETER VERHASSELT ET AL.: "Twelve open reading frames revealed in the 23.6kb segent flanking the centromere on the Saccharomyces cerevisiae chromosome XIV right arm"	1-23,27
·	YEÅST, vol. 10, no. 7, July 1994 (1994-07), pages 1355-1361, XP002112572 abstract; table 2	
X	-& Swissprot Database Entry Yn84_Yeast Accession number P40345; 1 February 1995 XP002112574 the whole document	1-23,27
	·	
	-/	

	LX.	Further documents are listed in the continuation of box C.	X
ı			

Patent family members are listed in annex.

- Special categories of cited documents:
- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or
- document published prior to the international filling date but later than the priority date claimed
- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled

Date of mailing of the international search report

"&" document member of the same patent family

Date of the actual completion of the international search

30/10/2000

17 October 2000

Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentiaan 2

NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl. Fax: (+31-70) 340-3016

Authorized officer

Montero Lopez, B

?

		FC1/EF 00/02/01	
	ation) DOCUMENTS CONSIDERED TO BE RELEVANT  Citation of document, with indication, where appropriate, of the relevant passages	- Relevant to	daim No
Category °	Citation of document, with indication, where appropriate, or the relevant passages	/ televalit (c	Classifi 140.
X	DATABASE EMBL 'Online! Database Entry SPBC776, 21 January 1999 (1999-01-21) LYNE M. ET AL.: "S. pombe chromosome II cosmid c776" Database accession no. AL035263 XP002150203 the whole document	1-	23,27
<b>X</b>	DATABASE EMBL 'Online! Database Entry AI398644, 10 February 1999 (1999-02-10) XP002150204 the whole document & MARY ANNE NELSON ET AL.: "Expressed sequences from conidial, mycelial, and sexual stages of Neurospora crassa "FUNGAL GENETICS AND BIOLOGY, vol. 21, 1997, pages 348-363, XP000952173	1-	23,27
<b>X</b>	KEITH STOBART ET AL.: "Triacylglycerols are synthesized and utilized by transacylation reactions in microsomal preparations of developing safflower (Carthamus tinctorius L.) seeds" PLANTA, vol. 203, no. 1, 1997, pages 58-66, XP002112573 page 58, right-hand column, last paragraph -page 59, left-hand column, paragraph 1 page 63, right-hand column, paragraph 2	25	
A	WO 98 55631 A (CALGENE LLC) 10 December 1998 (1998-12-10) page 9, line 36 -page 10, line 7 page 12, line 28 -page 13, line 18 page 14, line 34 -page 15, line 13 page 20, line 5 -page 25, line 4	1-	27
P,X	DATABASE SWALL 'Online! Database Entry 094680, 1 May 1999 (1999-05-01) LYNE M. ET AL.: "hypothetical 69.7 kDa protein C776.14 in chromosome II" Database accession no. 094680 XP002150205 the whole document	1-	23,27

## INTERNATIONAL SEARCH REPORT

rmation on patent family members

Internal al Application No PCT/EP 00/02701

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9855631 A	10-12-1998	CN 1266460 T EP 1003882 A	13-09-2000 31-05-2000

# This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

BLACK BORDERS
☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
☐ FADED TEXT OR DRAWING
☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
☐ SKEWED/SLANTED IMAGES
☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
☐ GRAY SCALE DOCUMENTS
LINES OR MARKS ON ORIGINAL DOCUMENT
☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
□ other.

## IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.